BMJ Open

HPV vaccine concern on Twitter: A cross-country evaluation of Australia, Canada, and the United Kingdom

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016869
Article Type:	Research
Date Submitted by the Author:	15-Mar-2017
Complete List of Authors:	Shapiro, Gilla; McGill University, Psychology Surian, Didi; University of New South Wales , Centre for Health Informatics Dunn, Adam; University of New South Wales Perry, Ryan; University of Melbourne, Centre for Health Policy Kelaher, Margaret; University of Melbourne, Centre for Health Policy
Primary Subject Heading:	Communication
Secondary Subject Heading:	Sexual health, Public health, Oncology, Infectious diseases, Health informatics
Keywords:	Human papillomavirus, Vaccination, Twitter, Health Belief Model, Social media

HPV vaccine concern on Twitter: A cross-country evaluation of Australia, Canada, and the United Kingdom

Gilla K. Shapiro (corresponding author), PhD Candidate, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada & Centre for Health Policy, School of Population and Global Health, University of Melbourne, Australia**;** Correspondence to: gilla.shapiro@mail.mcgill.ca

Didi Surian, Post-Doctoral Research Fellow, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 75 Talavera Road, Sydney, New South Wales, 2109, Australia

Adam Dunn, Senior Research Fellow**,** Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 75 Talavera Road, Sydney, New South Wales, 2109, Australia

Ryan Perry, Researcher, Centre for Health Policy, School of Population and Global Health, University of Melbourne, 235 Bouverie Street, Melbourne, Victoria, 3053, Australia

Margaret Kelaher, Professor, Head Evaluation and Implementation Science, Centre for Health Policy, School of Population and Global Health, University of Melbourne, 235 Bouverie Street, Melbourne, Victoria, 3053, Australia

Abstract

Objective: Opposition to HPV vaccination is common on social media and has the potential to impact vaccine coverage. This study aims to conduct an international comparison of the proportions of tweets about HPV vaccines that express concerns, the types of concerns expressed, and the social connections among users posting about HPV vaccines.

Design: An international comparison of English language tweets about HPV vaccines and social connections among Twitter users posting about HPV vaccines between January 2014 and April 2016 was conducted.

Setting: The content of tweets and the social connections between users who posted tweets about HPV vaccines from Australia, Canada and the United Kingdom.

Population: 16,789 Twitter users who posted 43,852 tweets about HPV vaccines.

Main outcome measures: The proportions of tweets expressing concern, the type of concern expressed, and the proportions of local and international social connections between users.

Results: Tweets expressing concerns about HPV vaccines made up 14.9% of tweets in Canada, 19.4% in Australia, and 22.6% in the UK. The types of concerns expressed were similar across the three countries, with concerns related to safety, pain, side effects, and logistical barriers being the most common. Users expressing concerns about HPV vaccines in each of the three countries had a relatively high proportion of international followers also expressing concerns.

Conclusions: The proportion and type of HPV vaccine concerns expressed on Twitter were similar across the three countries. Twitter users who mostly expressed concerns about HPV vaccines were better connected to international users who shared their concerns. The possibility of an international community of users who express vaccine concerns may influence the speed at which controversies and misinformation gain traction. International

coordination across public health organizations, proactive online monitoring of misinformation, and discussions with physicians may support timely and targeted media interventions to address vaccine concerns.

Key words: Human papillomavirus; Vaccination; Twitter; Health Belief Model; Social media.

Strengths and limitations of this study

- This study conducted an international comparison investigating how vaccination concerns are expressed on Twitter.
- Machine learning methods were used to identify and classify the proportion and types of concerns expressed in thousands of tweets.
- By collecting and analyzing social connections among Twitter users posting about HPV vaccines in three countries, the study revealed the potential for misinformation and concerns to spread internationally.
- While Twitter is used by a substantial number of people, it cannot be used to generalize about vaccine concerns held by the general population.
- This study used follower networks to examine social connections; however, future research should evaluate other possible ways of interacting on Twitter including 'liking' or 'retweeting' posts.

Introduction

Human papillomavirus (HPV) is a prevalent sexually transmitted infection that can cause cancers and anogenital warts.[1-5] Since 2006, three prophylactic vaccines have been developed to protect adolescents from HPV-associated health problems.[6] Research has demonstrated that these vaccines are safe and effective in reducing HPV related infections, genital warts, and pre-cancers.[7-12] As a result, at least sixty-five countries have implemented HPV vaccination programs for females in their national immunization schedules.[11] However, due to parental attitudes and concerns HPV vaccine coverage remains suboptimal, hindering cancer prevention efforts.[13]

The media has the potential to dramatically impact vaccine coverage through influencing parental awareness, perception, and attitudes.[14-18] Unconfirmed reports of adverse events associated with the HPV vaccine published in the media dramatically affected female HPV vaccine coverage in Japan and Columbia.[11,19,20] Many individuals use the internet and social media to access health information; however, these sources have been described as a risky platform that can rapidly amplify unbalanced, distorted or inaccurate information about vaccines.[14,21-23] For example, a study by Betsch et al. found that even 5 to 10 minutes of access to vaccine-critical websites negatively influenced individuals' risk perception and intentions to be vaccinated.[24] Similarly, Nan and Madden report that, compared to a control group, participants who were exposed to negative online blogs about HPV perceived the vaccine as less safe, held more negative attitudes, and reported a reduced intention to receive the vaccine.[25]

Previous research has evaluated the public discourse concerning HPV vaccination in newspapers,[26-29] online news,[30] comments to online news articles,[31] Facebook,[32] blogs or online forums,[25,33,34] and YouTube videos.[35,36] Twitter is an important source of information regarding HPV and vaccine hesitancy,[21,37,38] and several studies have examined the representation of HPV vaccines on Twitter.[22,32,39-43] Though many of these studies analyse a limited number of HPV-related tweets, a few have used data mining and machine learning techniques to analyse a large number of tweets.[40,41,43,44] However, no research has conducted an international comparison to evaluate and compare how vaccination concerns are expressed across countries. Furthermore, no research has examined the domestic and international network connectedness of HPV vaccine concern expression.

The aim of this study was to explore the proportion of HPV vaccine concern on Twitter, examine the type of concern expressed in Australia, Canada, and the United Kingdom (UK), and investigate differences in the ways Twitter users connected locally and internationally.

Methods

Study overview

Tweets related to HPV vaccines during were collected from January 2014 to April 2016 in Australia, Canada, and the UK. Data captured included information about users' locations, the text of the tweets, and information about social connections. To enable the classification of a large number of tweets, two stages of machine learning classifiers were constructed from a sample of tweets that were manually coded by two investigators.

Study data

Using a similar approach to previous studies that examined large number of tweets in communities of Twitter users posting about HPV vaccines,[43,45] the Twitter Search Application Programming Interface (API) was used to collect tweets in the English language about HPV vaccines from January 2014 to April, 2016. The search terms were "Gardasil", "Cervarix", "hpv AND vaccin*", and "cervical AND vaccin*". Information extracted from each tweet included the unique tweet identifier, tweet text, creation time, the identifier of the user posting the tweet, and geographical coordinates (if available). Without any restrictions applied to the locations of users, the entire dataset included 358,194 tweets (including retweets) by 129,286 users.

A gazetteer was used to transform the text provided by users into coordinates, and any users with self-reported locations that were located in coordinates in Australia, Canada, or the UK were included in this analysis (Supplemental Material, Section 1).

Other data that were used in the analyses included the set of social connections formed among the users who were included in the analyses. For each user, the Twitter Search API was used to collect the set of all follower relationships in which the user was involved, shortly after the first time the user posted a relevant tweet in the period. A network was then formed to include all users who tweeted about HPV vaccines from the three countries, and the follower relationships defined the social connections in an unweighted, directed network (Supplemental Material, Section 2).

The Macquarie University Human Research Ethics Committee (#5201401028) and the University of Melbourne's Research Ethics Board (#1647488.1) provided ethics approval for data collection and analysis.

Analysis

Supervised machine learning methods were used to classify the tweets in two stages; firstly to identify tweets that expressed any concern and secondly to classify specific types of concerns. In the first stage, 1000 tweets were sampled from the set of all tweets to manually label those that expressed concerns. In the second stage, 1000 tweets were sampled from the set of tweets that were estimated to be concerns to manually label them by type. The manually labelled tweets were used to train classifiers to label any tweet by the type of concern expressed (Supplemental Material, Section 3).

The categories for types of tweets expressing concerns were determined using an inductive and deductive procedure. The Health Belief Model, one of the most widely used theories in health psychology, was used as the basis for coding the types of HPV vaccine concerns expressed on Twitter.[46] The Health Belief Model has been used previously to evaluate the determinants of HPV vaccination and non-compliance by identifying perceived susceptibility to HPV, perceived severity of HPV, perceived benefits of HPV vaccination, perceived barriers of HPV vaccination (e.g. logistical barriers and potential physical costs/harms as a consequence of receiving the HPV vaccine), and cues to action (e.g. influences prompting HPV vaccine uptake such as information from health care provides, family, or friends).[36,46-48] To account for additional prominent concerns that were not captured by the model, the constructs were also informed by previous content analyses of media and social media related to the HPV vaccine, as well as literature on vaccine hesitancy.[17,22,26,30,31,34,36,49-58] The coding scheme was therefore extended to include mistrust, undermining of religious principles, undermining of civil liberties, additional concerns (not otherwise specified), and ambiguous tweets (Table 1). The coding scheme was used by two investigators (GS and RP) to code 12 types of concerns expressed in a second sample of 1000 tweets. After examining the proportions of different types of concerns in the sample and accuracy of the multi-class classifier (Supplemental Material, Sections 3 and 4), types of concerns were combined to improve the performance that could be achieved by the machine learning classifiers (Table 1).

The network of social connections formed by users who posted tweets about HPV vaccines was used to compare the proportions of local (within a country) and international (across countries) followers. The group of users for whom at least half of their relevant tweets were expressing concerns were assigned to one group (concern), and all other users were assigned to another (non-concern). The users were then also split by country, and the proportions of local and international followers were compared across groups.

Results

There were 129,286 Twitter users who posted at least one tweet about HPV vaccines during the period. The location inference method identified 2,792 (2.2%) of those users located in Australia, 7,237 (5.6%) located in Canada, and 6,760 (5.2%) users located in the UK (Table 2).

From the 16,789 users in the three countries, a total of 43,852 tweets about HPV vaccines were posted, of which 7,173 (16.4%) were from Australia, 18,927 (43.2%) were from Canada, and 17,752 (40.5%) were from the UK. This corresponded to an average of 2.57 tweets per user in Australia (range, 1-198), 2.61 tweets per user in Canada (range, 1-433), and 2.62 tweets per user in the UK (range, 1-501).

Expressions of concern

When labelling tweets that expressed concerns, the binary classifier (stage one) achieved a recall of 0.97 and a precision of 0.90. This indicates that the binary classifier missed 3% of tweets that were manually labelled as having expressed a concern and 10% of tweets it labelled as having expressed a concern were manually labelled otherwise. Because the multi-class classifier (at stage two) identified a proportion of these mislabelled tweets in the second round, the overall rate of error in stage one was within 5% of the correct proportion (Supplemental Material, Section 4).

The proportion of tweets posted about HPV vaccines from users in the three countries expressing concerns was 18.7% (8,215 of 43,852 tweets), but there were differences in these proportions across the three countries (Table 2). Canada had the lowest proportion of tweets expressing concerns at 14.9% (2,818 of 18,927 tweets), followed by Australia at 19.3% (1,388 of 7,173 tweets). The UK had the highest proportion of tweets expressing concerns at 22.6% (4,009 of 17,752 tweets). Tweets expressing concerns also tended to have smaller audiences compared with tweets not expressing concern about HPV vaccines (Supplemental Material, Section 3).

Types of concerns expressed

When identifying concerns related to cues to action the classifier respectively produced a precision of 0.81 and a recall of 0.74. For barriers, the precision was 0.91 and recall was 0.92. The classifier was less reliable for the remainder of the concern groups because these types of concerns made up a much smaller proportions, resulting in imbalance in the data, which affects the performance that can be achieved by the classifiers (Supplemental Material, Section 5).

Tweets expressing concerns about barriers comprised the largest type of concern by both the proportion of tweets expressing concerns (Table 3). The proportions of each group of concerns across the three countries were generally consistent.

Social connections among users

Among users from the three countries who posted about HPV vaccines, 18.2% (3,062 of 16,789) were labelled as having expressed concerns (at least half of the tweets about HPV vaccines they posted were labelled as having expressed a concern). The total number of follower connections among the set of 16,789 users was 502,629. Users from the three countries were disproportionately more likely to be followed by users from the same country, creating clusters of users by country (Figure 1).

Relative to users who did not express concern about HPV, users that did express concerns had a higher proportion of international followers who also expressed concerns (Figure 2). Among UK users expressing concerns, 26.1% of followers also expressed concerns (compared to 9.1% of followers among UK users not expressing concerns). Also among UK users expressing concerns, 28.0% of their followers also expressed concerns and were from Australia or Canada, and 9.9% of their followers did not express concerns and were from Australia or Canada. In comparison, among UK users not expressing concerns, only 5.8% of their followers were users not expressing concerns and from Australia or Canada, and 8.3% of their followers were users expressing concerns and from Australia or Canada (Supplemental Material, Section 6). This pattern was consistent across each of the three countries. The results indicate that users who mostly expressed concerns were disproportionately well-connected to international users discussing HPV vaccines.

Discussion

This study found that in Australia, Canada, and the UK nearly one in five of the tweets about HPV vaccines were expressions of concern. Canadian Twitter users less often expressed concerns about HPV vaccines compared to Australia and the UK. There was a general consistency in the proportions of specific concerns across the three countries, and the most common concerns were related to safety, pain, side effects, and logistical barriers (such as accessibility and affordability). The results demonstrated that users expressing concerns about HPV vaccines tended to be relatively well connected to users discussing HPV vaccine concerns in other countries, especially between Canada and the UK.

Previous studies examining the representation of HPV vaccines on Twitter identified slightly higher proportions of negative tweets or tweets expressing concerns, but these studies captured different time periods and did not compare specific countries.[41,45] In examining the type of concern expressed about the HPV vaccine on other social media sites, researchers have also observed the predominance of 'barrier' concerns (i.e. safety, pain, side effects, logistical barriers).[33,34,59] However, while the present research study found concerns about safety were most common on Twitter; other research found safety to be surpassed or similar in salience to other prevalent themes including conspiracies/search for truth, mistrust for health system, and promoting promiscuity.[33,34,59] Surian et al. analysed topics regarding HPV vaccines on Twitter and found individuals who posted about 'harms and conspiracies' posted more often than other users, suggesting that some users are actively seeking to introduce concerns about HPV vaccines into the public domain.[43] The predominance of 'barrier' concerns on Twitter indicates the importance of physicians discussing concerns about vaccine safety with their patients.

International networking on twitter suggests that vaccine related controversies in one country could reverberate around the world and impact vaccine coverage. Public health professionals and policymakers must therefore be able to monitor, rapidly identify, and react to such concerns (e.g. by providing evidence-based responses in real-time and strengthening their own international networks).[23,39,60] This research provides public health practitioners and policymakers with evidence that concerns about 'barriers' on Twitter are widespread; effective communication campaigns could be designed and implemented to target this concern in locations where it is likely to have the greatest impact. However, it is important for future research to design and evaluate appropriate messaging of such a campaign so that this intervention does not 'backfire' and increase hesitancy.[61]

This study also found that Twitter users propagating HPV vaccine concerns tended to have higher proportions of international connections compared to those not expressing concerns. It may be helpful for pro-HPV vaccine groups (such as public health organizations) to monitor emerging concerns about vaccines occurring internationally on social media in order to pre-empt and respond to misinformation locally. Such support could have been beneficial for Japan and Colombia during their media-influenced dramatic decrease in HPV vaccine coverage.[11,19,20]

Further research in this area should consider the relationship between the information about HPV vaccines that enters into the public discourse and the decision-making of individuals and populations. While Canada had the lowest proportion of tweets in which concerns were expressed of the three countries, it also has the lowest rate of HPV vaccine uptake.[49,62-65] Accordingly, we echo Gollust et al.'s recent call for greater experimental research designs to make causal assertions about the impact of the media on vaccine coverage.[66] Although some studies have begun to do so,[25,67] it would be helpful for future research to specifically evaluate the impact of Twitter messages and for moderating variables to also be evaluated.

There were several limitations to this study. First, the findings are specific to Twitter, and while Twitter represents one of the largest populations of social media users, the results are not necessarily representative of the broader public discourse about HPV vaccines in news and online social media.[68-71] Second, while the location inference method is a standard in the area,[72,73] the methods are imperfect.[74-77] Finally, using networks based on which users follow each other does not necessarily capture all of the interactions that occur online. While some argue that interaction with content is a better measure of impact than followers,[78] others have argued that many users on Twitter are passive and do not interact with the content,[79] and as such followers may be a better indicator of impact. As this study examined follower networks, it would be helpful for future research to compare followers to different ways of interacting with content in order to better understand the impact of HPV vaccine tweets.

Conclusions

This study characterized the concerns about HPV vaccines expressed by Twitter users in three countries. The UK had the greatest proportion of tweets expressing concerns about HPV vaccines and Canada had the least, and the types of concerns expressed were relatively consistent across the three countries. Users who expressed concerns about HPV vaccines were generally more closely connected to users in other countries who also expressed concerns, suggesting that controversies and misinformation may be rapidly shared across international boundaries. This research could be used to design public health interventions that address concerns about the HPV vaccines on Twitter, and suggests that methods for addressing vaccine concerns may benefit from physician discussions about vaccine safety and further coordination of public health agencies internationally.

Acknowledgement

Contributors GS, AD, and MK conceptualized and designed the study. DS and AD acquired the data. GS drafted the manuscript. All authors analysed and interpreted the data, critically revised the manuscript for important intellectual content, and approved the final version.

Funding: GS was supported by the CIHR Vanier Canada Graduate Scholarship and Queen Elizabeth II Diamond Jubilee Scholarship programs. AD is an investigator on a related NHMRC Project Grant (APP1128968). The funding source had no role in the design of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

Competing interests: None.

Ethics approval: Macquarie University Human Research Ethics Committee (#5201401028) and the University of Melbourne's Research Ethics Board (#1647488.1).

Data sharing statement: No additional data is available. Due to ethical approvals and the Twitter Terms of Service, the individual identifiers for Twitter posts and users cannot be provided with their manual or automatic labels.

References

1. Trottier H, Franco EL. The epidemiology of genital human papillomavirus infection. *Vaccine.* 2006;24(1):S1–15.

2. Tota JE, Chevarie-Davis M, Richardson LA, Devries M, Franco EL. Epidemiology and burden of HPV infection and related diseases: implications for prevention strategies. *Prev Med* 2011;53(1):S12-21.

3. Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *The Lancet Global Health.* 2016;4(9):e609–e616

4. Forman D, de Martel C, Lacey CJ, et al. Global burden of human papillomavirus and related diseases. *Vaccine.* Nov 20 2012;30 Suppl 5:F12-23.

5. National Advisory Committee on Immunization. Update On Human Papillomavirus Vaccines. *Canada Communicable Disease Report.* 2012;38.

6. Canadian Cancer Society. *Canadian Cancer Statistics 2016.* 2016.

7. Markowitz LE, Dunne EF, Saraiya M, et al. Human papillomavirus vaccination: recommendations of the Advisory Committee on Immunization Practices. *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control.* Aug 29 2014;63(Rr-05):1-30.

8. Ferris D, Samakoses R, Block SL, et al. Long-term study of a quadrivalent human papillomavirus vaccine. *Pediatrics.* Sep 2014;134(3):e657-665.

9. Garland SM. The Australian experience with the human papillomavirus vaccine. *Clin Ther* 2014;36:17-23.

10. Brotherton JM, Saville AM, May CL, Chappell G, Gertig DM. Human papillomavirus vaccination is changing the epidemiology of high-grade cervical lesions in Australia. *Cancer Causes Control.* 2015;26(6):953–954.

11. Brotherton JML, Bloem PJN. HPV Vaccination: Current Global Status. *Curr Obstet Gynecol Rep.* 2015;4:220-233.

12. Drolet M, Benard E, Boily MC, et al. Population-level impact and herd effects following human papillomavirus vaccination programmes: A systematic review and meta-analysis *Lancet Infect Dis.* 2015;15(5):565-580.

13. Holman DM, Benard V, Roland KB, Watson M, Liddon N, Stokley S. Barriers to Human Papillomavirus Vaccination Among US Adolescents: A Systematic Review of the Literature. *JAMA pediatrics.* 2014; 168 (1):76-82.

14. Kumar D, Chandra R, Mathur M, Samdariya S, Kapoor N. Vaccine hesitancy: Understanding better to address better. *Isr J Health Policy Res.* 2016;5(2).

15. Mason BW, Donnelly PD. Impact of a local newspaper campaign on the uptake of the measles mumps and rubella vaccine. *J Epidemiol Community Health.* 2000;54:473-474.

16. Gangarosa EJ, Galazka AM, Wolfe CR, et al. Impact of anti-vaccine movements on pertussis control: the untold story. *Lancet.* 1998;351(9099):356-361.

17. Larson HJ, Jarrett C, Schulz WS, et al. Measuring vaccine hesitancy: The development of a survey tool. *Vaccine.* 2015.

18. Gollust SE, Attanasio L, Dempsey A, Benson AM, Fowler EF. Political and news media factors shaping public awareness of the HPV vaccine. *Womens Health Issues.* 2013;23(3):e143-151.

19. Hanley SJ, Yoshioka E, Ito Y, Kishi R. HPV vaccination crisis in Japan. *Lancet.* 2015;385(9987):2571.

20. El Tiempo. Caso de niñas del Carmen de Bolívar desplomó vacunación contra el VPH. February 14, 2015; http://www.eltiempo.com/estilo-de-vida/salud/vacuna-contra-el-papiloma-humano-cayo-58-puntos-porcentuales/15246061.

21. Gabarron E, Serrano JA, Wynn R, Lau AYS. Tweet content related to sexually transmitted diseases: no joking matter. *Journal of medical Internet research.* 2014;16(10):e228.

22. Britt RK, Hatten KN, Chappuis SO. Perceived behavioral control, intention to get vaccinated, and usage of online information about the human papillomavirus vaccine. *Health psychology and behavioral medicine.* 2014;2(1):52-65.

23. Amicizia D, Domnich A, Gasparini R, Bragazzi NL, Lai PL, Panatto D. An overview of current and potential use of information and communication technologies for immunization promotion among adolescents. *Human Vaccines & Immunotherapeutics* 2013;9(12):2634–2642.

24. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. *J Health Psychol.* 2010;15:446–455.

25. Nan X, Madden K. HPV vaccine information in the blogosphere: how positive and negative blogs influence vaccine-related risk perceptions, attitudes, and behavioral intentions. *Health communication.* 2012;27(8):829-836.

26. Perez S, Fedouruk C, Shapiro GK, Rosberger Z. Giving Boys a Shot: The HPV Vaccine's Portrayal in Canadian Newspapers. *Health communication.* 2016;31(12):1527-1538.

27. Robbins SC, Pang C, Leask J. Australian newspaper coverage of human papillomavirus vaccination, October 2006-December 2009. *Journal of health communication.* 2012;17(2):149-159.

28. Krieger JL, Katz ML, Eisenberg D, Heaner S, Sarge M, Jain P. Media coverage of cervical cancer and the HPV vaccine: implications for geographic health inequities. *Health expectations : an international journal of public participation in health care and health policy.* Sep 2013;16(3):e1-12.

29. Quintero JJ, Sionean C, Scott AM. Exploring the presentation of news information about the HPV vaccine: a content analysis of a representative sample of U.S. newspaper articles. *Health communication.* 2011;26(6):491-501.

30. Habel MA, Liddon N, Stryker JE. The HPV vaccine: A content analysis of online news stories. *J Womens Health.* 2009;18(3):401-407.

31. Feinberg Y, Pereira JA, Quach S, et al. Understanding Public Perceptions of the HPV Vaccination Based on Online Comments to Canadian News Articles. *PloS one.* 2015;10(6):e0129587.

32. Coloma PM, Becker B, Sturkenboom MC, van Mulligen EM, Kors JA. Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies. *Drug safety.* Oct 2015;38(10):921-930.

33. Penta MA, Baban A. Dangerous agent or saviour? HPV vaccine representations on online discussion forums in Romania. *International journal of behavioral medicine.* 2014;21(1):20-28.

34. Keelan J, Pavri V, Balakrishnan R, Wilson K. An analysis of the Human Papilloma Virus vaccine debate on MySpace blogs. *Vaccine.* Feb 10 2010;28(6):1535-1540.

35. Ache KA, Wallace LS. Human Papillomavirus vaccination coverage on YouTube. *Am J Prev Med* 2008;35(4):389-392.

36. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. *Health communication.* 2012;27:478–485.

37. Gunasekaran B, Jayasinghe Y, Fenner Y, et al. Knowledge of human papillomavirus and cervical cancer among young women recruited using a social networking site. *Sexually transmitted infections.* 2013;89(4):327-329.

38. McNeill A, Harris PR, Briggs P. Twitter Influence on UK Vaccination and Antiviral Uptake during the 2009 H1N1 Pandemic. *Frontiers in public health.* 2016;4:26.

39. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly available online tool facilitates real-time monitoring of vaccine conversations and sentiments. *Health Affairs.* 2016;35(2):341-347.

40. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *Journal of medical Internet research.* 2015;17(6):e144.

41. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in health technology and informatics.* 2015;216:761-765.

42. Mahoney LM, Tang T, Ji K, Ulrich-Schad J. The Digital Distribution of Public Health News Surrounding the Human Papillomavirus Vaccination: A Longitudinal Infodemiology Study. *JMIR Public Health Surveill.* 2015;1(39):e2.

43. Surian D, Nguyen QD, Kennedy G, Johnson M, Coiera E, Dunn GA. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *Journal of medical Internet research.* 2016;18(8):e232.

44. Salathé M, Vu DQ, Khandelwal S, Hunter DR. The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science.* 2013.

45. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on

Social Media: An Observational Study. *Journal of Medical Internet Research.* 2015;17(6):e144.

46. Rosenstock IM. Historical origins of the health belief model *Health Educ Monogr* 1974;2:328-335.

47. Donadiki EM, Jiménez-García R, Hernández-Barrera V, et al. Health Belief Model applied to non-compliance with HPV vaccine among female university students. *Public Health.* 2014;128(3):268-273.

48. Krawczyk A, Knäuper B, Gilca V, et al. Parents' decision-making about the human papillomavirus vaccine for their daughters: I. Quantitative results. *Hum Vaccin Immunother.* 2015;11(2):322-329.

49. Perez S, Shapiro GK, Brown CA, Dube E, Ogilvie G, Rosberger Z. 'I didn't even know boys could get the vaccine': Parents' reasons for human papillomavirus (HPV) vaccination decision making for their sons. *Psycho-Oncology.* 2015;[Epub ahead of print].

50. Larson H, Leask J, Aggett S, Sevdalis N, Thomson A. A Multidisciplinary Research Agenda for Understanding Vaccine-Related Decisions. *Vaccines.* 2013;1(3):293-304.

51. Casciotti DM, Smith KC, Tsui A, Klassen AC. Discussions of adolescent sexuality in news media coverage of the HPV vaccine. *Journal of adolescence.* 2014;37(2):133-143.

52. Casciotti DM, Smith KC, Andon L, Vernick J, Tsui A, Klassen AC. Print news coverage of school-based human papillomavirus vaccine mandates. *The Journal of school health.* 2014;84(2):71-81.

53. Madden K, Nan X, Briones R, Waks L. Sorting through search results: a content analysis of HPV vaccine information online. *Vaccine.* 2012;30(25):3741-3746.

54. Nan X, Daily K. Biased assimilation and need for closure: Examining the effects of mixed blogs on vaccine-related beliefs. *Journal of health communication.* 2015;20(4):462-471.

55. Thomson A, Robinson K, Vallee-Tourangeau G. The 5As: A practical taxonomy for the determinants of vaccine uptake. *Vaccine.* 2016;34(8):1018-1024.

56. Wolfe RM, Sharp LK, Lipsky MS. Content and design attributes of antivaccination web sites. *JAMA.* 2002;287(24):3245-3248.

57. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: A content analysis. *JAMA.* 2007;298(21):2482-2484.

58. Abdelmutti N, Hoffman-Goetz L. Risk messages about HPV, cervical cancer, and the HPV vaccine Gardasil: A content analysis of Canadian and U.S. national newspaper articles. *Women & health.* Jul-Aug 2009;49(5):422-440.

59. Kata A. A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine.* 2010;28:1709-1716.

60. Larson HJ, Wilson R, Hanley S, Parys A, Paterson P. Tracking the global spread of vaccine sentiments: The global response to Japan's suspension of its HPV vaccine recommendation. *Hum Vaccin Immunother.* 2014;10(9):2543-2550.

61. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: A randomized trial. *Pediatrics.* 2014;133:e835-e842.

62. National HPV Vaccination Program Register. *National (Australia) HPV 3 dose vaccination coverage for females turning 15 years of age in 2015.* 2016.

63. National HPV Vaccination Program Register. National (Australia) HPV 3 dose vaccination coverage for males turning 15 years of age in 2015. 2016.

64. Gilbert NL, Gilmour H, Dubé E, Wilson SE, Laroche J. Estimates and determinants of HPV non-vaccination and vaccine refusal in girls 12 to 14 years of age in Canada:

Results from the Childhood National Immunization Coverage Survey, 2013. *Human Vaccines & Immunotherapeutics.* 2016;12(6):1484-1490.

65. Vishram B, Byrne L, White J, Edelstein M. *Human Papillomavirus (HPV) vaccination coverage in adolescent females in England: 2014/15.* London, England: Public Health England;2015.

66. Gollust SE, LoRusso SM, Nagler RH, Fowlerc EF. Understanding the role of the news media in HPV vaccine uptake in the United States: Synthesis and commentary. *Hum Vaccin Immunother.* 2016;12(6):1430-1434.

67. Betsch C. Social media targeting of health messages: A promising approach for research and practice. *Hum Vaccin Immunother.* 2014;10(9):2636-2637.

68. Weeg C, Schwartz HA, Hill S, Merchant RM, Arango C, Ungar L. Using Twitter to Measure Public Discussion of Diseases: A Case Study. *JMIR Public Health Surveill.* 2015;1(1):e6.

69. Pew Research Center. *Social networking fact sheet.* http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/.

70. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in twitter: A quantitative analysis. *JMIR Public Health Surveill.* 2016;2(1):e1.

71. Paul MJ, Dredze M. You are what you tweet: Analyzing twitter for public health. *Association for the Advancement of Artificial Intelligence* 2011.

72. Kumar S, Morstatter F, Marshall G, Liu H, Nambiar U. Navigating Information Facets on Twitter (NIF-T). Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD); 2012.

73. Kounadi O, Lampoltshammer TJ, Groff E, Sitko I, Leitner M. Exploring twitter to analyze the public's reaction patterns to recently reported homicides in London. *PloS one.* 2015;10(3):e0121848.

74. Elmongui HG, Morsy H, Mansour R. Inference models for twitter user's home location prediction. IEEE/ACS International Conference of Computer Systems and Applications (AICCSA); 2015.

75. Bo H, Paul C, Timothy B. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research.* 2014;49:451-500.

76. Jurgens D. That's what friends are for: Inferring location in online social media platforms based on social relationships. . Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM); 2013.

77. Cheng Z, Caverlee J, Lee K. You are where you tweet: A content-based approach to geo-locating Twitter users. Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM); 2010.

78. Cha M, Haddadi H, Benevenuto F, Gummadi KP. Measuring user influence in Twitter: The million follower fallacy. Association for the Advancement of Artificial Intelligence 2010.

79. Romero DM, Galuba W, Asur S, Huberman BA. Influence and passivity in social media. In: Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, Michalis Vazirgiannis, eds. *Machine Learning and Knowledge Discovery in Databases* 2011:18-33.

Tables and Figures

Table 1. Coding scheme for the types of concerns expressed on Twitter

Original 12 types of concern	Combined concern group label	Example tweet
Not beneficial: stating that the HPV vaccine is not beneficial or useful (HBM construct 'perceived benefits')	Unnecessary	"Breaking Report HPV Cancers Rising In Spite of Vaccination URL #vaccination #guardasil #cervarix #HPV #cancer #fraud"
Logistical barriers: stating logistical barriers such accessible or affordability challenges (HBM construct 'perceived barriers')	Barriers	"Makes no sense that girls are covered for the HPV vaccine but I gotta pay $400 for it... #needstochange"
Physical harms: stating concerns about the physical issues or harms as a result of receiving the HPV vaccine including pain, safety, or side effects (HBM construct 'perceived barriers')	Barriers	"Mum still reckons im unwell like this all the time cos of the hpv vaccine. Thinking she might be right. Hpv vacc has loads of side effects"
Not severe: stating that HPV and/or its consequences are not severe (e.g. because it is common or clears up on its own) (HBM construct 'perceived severity')	Unnecessary	"The New Gardasil Is It Right For Your Daughter URL"
Low susceptibility: stating the HPV vaccine is unnecessary because there is a low likelihood of getting HPV and/or its consequences (HBM construct 'perceived susceptibility')	Unnecessary	"30 Facts you probably don\u2019t know about HPV and Gardasil...URL"
Cues to action: stating the influence of significant others guiding against receiving HPV vaccination (HBM construct 'cues to action')	Cues to action	"American College of Pediatricians warns about toxic effects of Gardasil vaccine"
Mistrust: stating a lack of confidence, mistrust, scepticism or belief in a HPV vaccine conspiracy	Additional concern	"Save dosh on the pharmaceutically lucrative, dubious Gardasil vaccine #qanda"
Undermining religious principles: stating concern that the HPV vaccine is inconsistent with religious principles	Additional concern	"…b. c. bishop, says chastity, not hpv vaccine, will keep girls healthy…"
Undermining civil liberties: stating concern about civil liberties (e.g. girls-only mandate, autonomy, who should be the decision maker for child vaccination, not being adequately consulted etc.)	Additional concern	"…had one dose of the gardasil at 17 after being bullied into it by my doctor, he basically told me I wasn't leaving without it"
Additional concerns not otherwise specified (e.g. belief in alternative medicine)	Additional concern	"…Our body does not need something NOT natural in our body to heal! The Gardasil/Vaccines were all in the..."
Tweet is ambiguous	Ambiguous	"…oh well if it's peer reviewed I'll give my son a gardasil shot."
No concern expressed	Non-concern	"Just saw a commercial that was like ask your doctor about Gardasil and I pumped my fist and shouted already did! because #sexualhealth"

Table 2. The total number of users and tweets from Australia, Canada, and the UK

Country	Number of users	Number of tweets	Tweets expressing concern	Tweets not expressing concern
Australia (%)	2,792 (16.6%)	7,173 (16.4%)	1,388 (19.4%)	5,785 (80.6%)
Canada (%)	7,237 (43.1%)	18,927 (43.2%)	2,818 (14.9%)	16,109 (85.1%)
UK (%)	6,760 (40.3%)	17,752 (40.5%)	4,009 (22.6%)	13,743 (77.4%)
Total	16,789 (100%)	43,852 (100%)	8,215 (18.7%)	35,637 (81.3%)

Table 3. Number of tweets, by country and concern type

Group label	Australia (%)	Canada (%)	UK (%)	Total (%)
Unnecessary	6 (0.39%)	13 (0.4%)	29 (0.6%)	48 (0.5%)
Barriers	717 (47.08%)	1,368 (42.4%)	2,137 (48.0%)	4,222 (45.9%)
Cues to Action	157 (10.31%)	274 (8.5%)	306 (6.9%)	737 (8.0%)
Additional concerns	187 (12.28%)	469 (14.5%)	560 (12.6%)	1,216 (13.2%)
Ambiguous	321 (21.08%)	694 (21.5%)	977 (22.0%)	1,992 (21.7%)
Total concern	1,388 (100%)	2,818 (100 %)	4,009 (100%)	8,215 (100%)

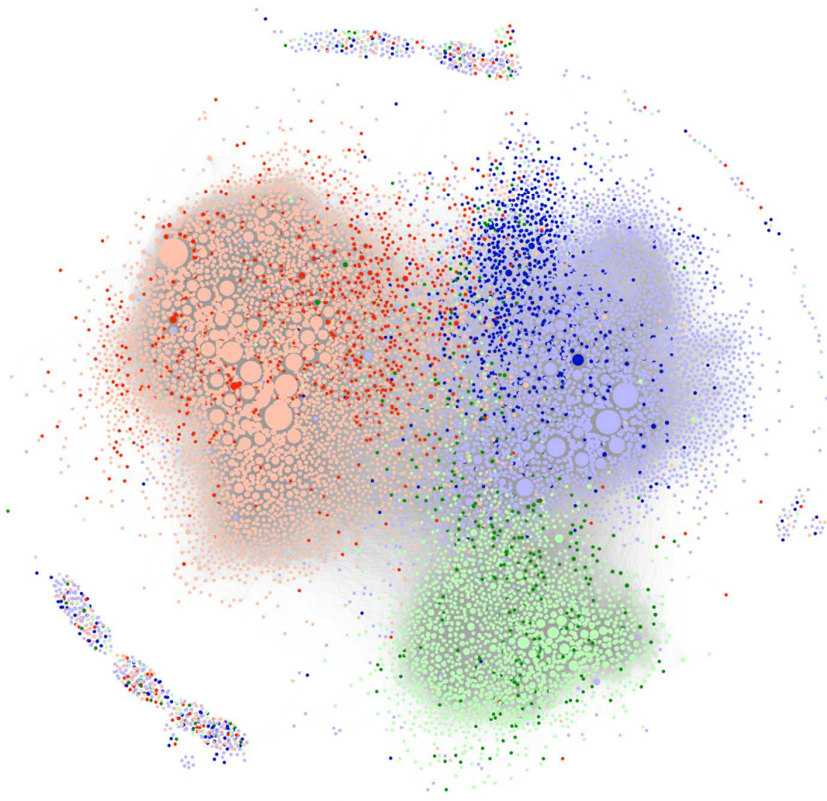


Figure 1. The follower network for Twitter users posting about HPV vaccines is coloured by country (Australia, green; Canada, red; UK, blue). Each node represents a user, and the node sizes are proportional to the number of followers within the user's network. Nodes are positioned by heuristic to be closer to nodes with which they are better connected, as a way of illustrating the community structure. Darker coloured nodes indicate users for whom at least 50% of their relevant tweets expressed concerns.

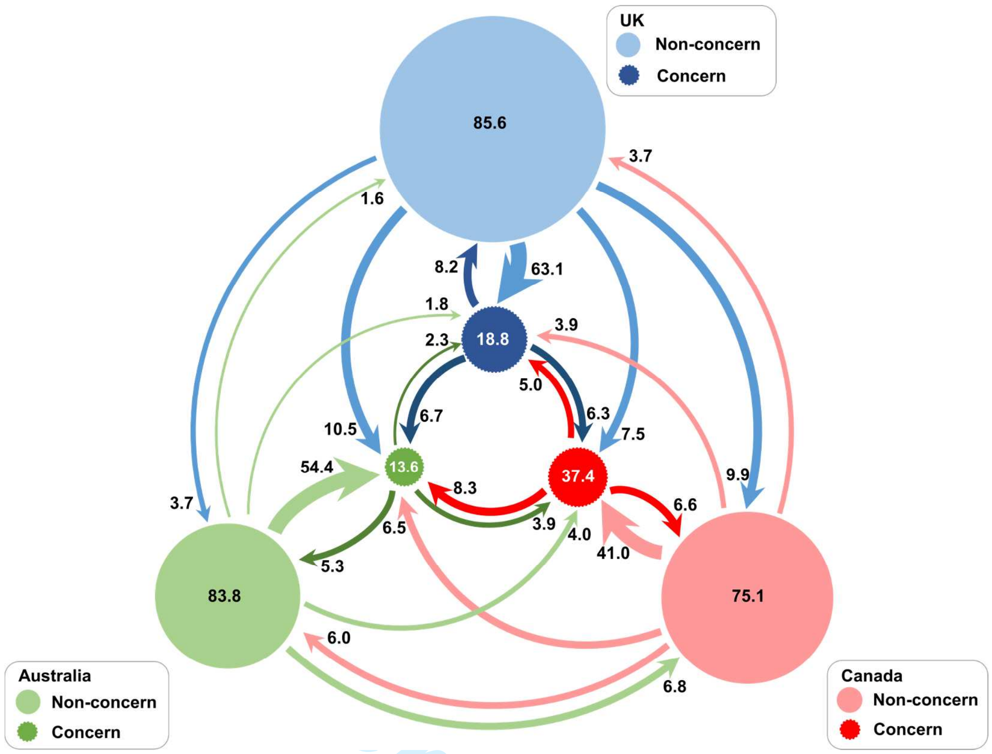


Figure 2. The percentages of followers for all users by expression of concern from Australia, Canada and the UK. The circle represents a concern group of Twitter users, where the circle size is proportional to the number of users. The arrow represents user following direction. The number represents the percentage of followers, where the number in a circle represents the percentage of followers from the same concern group. Only values above 1.5% are shown, see Supplemental Material (Section 5) for all values.

Supplemental Material

The information described below provides greater detail of the methodology and results described in the manuscript.

1. Extracting and estimating the home locations of Twitter users

The use of 'geo-tags' on Twitter—where geographical coordinates are embedded in the metadata—is relatively rare and unevenly distributed, which makes geo-tags an unreliable way to determine the locations of users.[1] A common alternative is to use the locations that users self-report in their profiles (free text).[2-4] Among the set of tweets collected for this study, 0.5% (1,735 of 358,194 tweets) included geo-tag information, while 70.1% (90,658 of 129,286 users) had some self-reported information about location in their user profiles.

Location inference methods were used to identify users located in Australia, Canada and United Kingdom. Nominatim,[5] a gazetteer, was used to translate the locations of Twitter users to identify users in the three countries. This information was taken from the tweet metadata (geo-tags) or user profile information (free text). Pre-processing steps for the user profile information included the removal of punctuation, numeric values, characters for non-English languages, and one-character words. Nominatim produces a score for the set of possible locations it returns, and a score of 0.4 was used as a threshold to avoid locations likely to be spurious (this threshold was determined through experiments in previous work). Where users included geo-tag information in the tweets they posted, the most frequent location was chosen (or the earliest where there were equally frequent and different locations used). Where users did not include geo-tag information, they were assigned to the location produced by Nominatim based on their user profile information.

2. Construction of the follower network

The social connections among the set of 16,789 users who posted about HPV vaccines and were located within Australia, Canada, and the UK were examined. The follower connections to and from each of the 16,789 users were collected through calls to the Twitter Application Program Interface (API), performed shortly after the first time each users posted a relevant tweet during the relevant time period. These data were used to construct the internal follower network by reconciling connections to and from each user to any other user in the set.

This study evaluated the proportions of international connections across the three countries, and examined the differences in the proportions of users who mostly post tweets expressing concerns about HPV vaccines relative to all other users. To do this, the ratios of follower connections of two types of users (those who express concerns in at least half of their relevant tweets versus all other users) in the three countries (Australia, Canada, and the UK), were compared.

3. Machine learning methods used to train and test the classification of the tweets

3.1. Pre-processing

The tweet texts were processed to construct features for our classifiers. No modifications were made on words that were hashtags (beginning with "#") and Twitter usernames (beginning with "@"). If a tweet text contained a website link (URLs), the domain name was stored. Standard data pre-processing including the removal of common English words,[6,7] and the removal of plurals and modifiers using Porter algorithm.[8] All numerical values were removed and all words were converted into lowercase. Each tweet was then transformed into a binary representation—a vector of length equal to the total number of unique features found across all tweets—with 1 marked for any feature in the tweet and 0 for all other features.

3.2. Supervised Machine Learning Training

Due to the large number of tweets collected in the period, a supervised machine learning approach was used to classify the tweets. This involved the manual labeling of a random sample of tweets, which were then used to train algorithms to identify similar patterns in the remaining tweets.[9,10] The data were classified in two stages to firstly identify tweets expressing concerns, and then to classify those concerns by type (Figure A1).

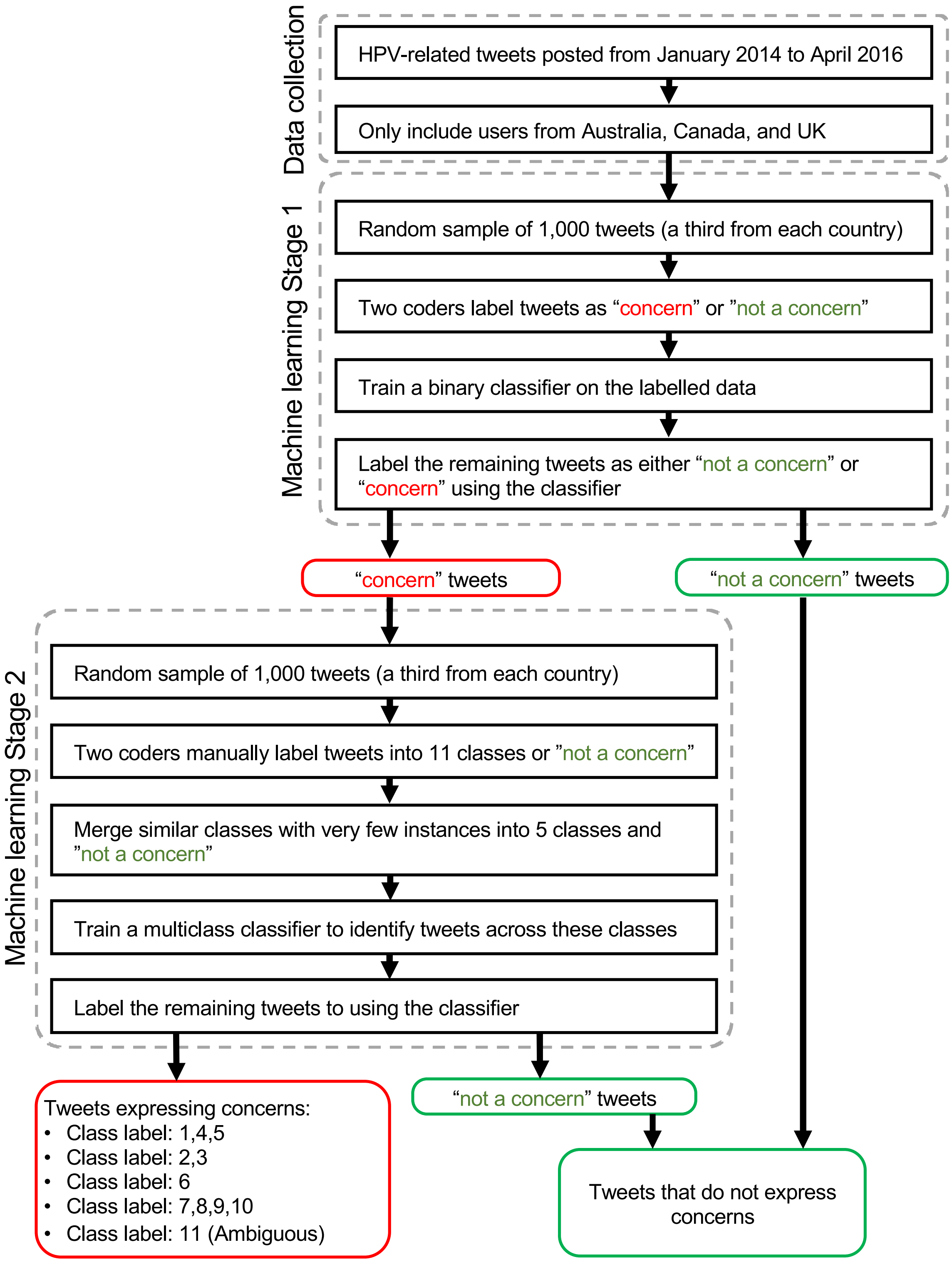


Figure A1. The design of the two-stage method for identifying classes of concerns in HPV vaccine tweets, where Stage 1 is the construction of a binary classifier and Stage 2 is the construction of a multiclass classifier

Classifiers for identifying any concern

The first stage of coding aimed to distinguish between tweets about HPV vaccines that expressed a concern versus tweets not expressing a concern. A random sample of 1,000 tweets were manually labeled as 'concern' or 'non-concern' by two investigators to form a training set from which to train a machine learning classifier. A separate set of 150 tweets was presented to the coders as a practice set for discussion prior to the independent labeling process. If the tone of the tweet was not immediately clear from the information provided within the tweet, the coders used links to webpages if they were available. There was strong agreement amongst the coders (92.3% agreement, Cohen's κ=0.81; 95% CI 0.77-0.85), and disagreements were resolved by discussion to produce the final training set.

A supervised binary classifier was trained using the manually labeled tweets to assign labels to the rest of the tweets in the data set. This study used a linear support vector machine (SVM) with stochastic gradient descent learning method to perform binary classification. The SVM method has been used widely for applications that deal with unbalanced and high dimensional data sets like those described here.[11-15] In this study, a random sample of 80% of the manually labeled tweets (the *training set*) was used to train and validate the classifier and the remaining 20% was used to test the performance of the classifiers (the *testing set*). The best parameters for the classifier were chosen using 10-fold cross validation using the training set. *K*-fold cross validation is a common method used to train a classifier in a prediction problem, where the training set is partitioned into *K* equal sized subsamples.[16] During the training process, the classifier is trained using all but one of the subsamples and validated on the remaining subsample, repeating the process *K* times. To avoid overfitting, the L2 regularization was used with 1,000 iterations during the training.[17,18]

Classifiers for identifying specific concerns

The second stage of coding aimed to distinguish different types of concern. The tweets classified as having expressed a concern about HPV vaccines in stage one were examined to distinguish the specific types of concerns. A random sample of 1,000 tweets was selected and a separate set of 150 tweets was used to pilot and test the scheme prior to the independent labeling process.

The categories for types of tweets expressing concerns were determined using an inductive and deductive procedure. Accordingly, the Health Belief Model (HBM) and additional concerns towards the HPV vaccine that have been identified in the literature were used to develop an initial coding scheme of 12 types of concerns (Table A1). These coding categories were discussed and agreed upon amongst the research team. There was good agreement in coding the random sample of 1,000 tweets (79.0% agreement, Cohen's κ=0.71; 95% CI 0.67-0.74), and any disagreements were resolved by discussion to produce the final training set. For example, the HBM factor of 'self-efficacy' was originally included in this coding scheme but deleted after team consultation as it overlapped with other groups during rounds of practice coding (i.e. 'logistical barriers').

Table A1. Number of coded tweets of each type of concern

Class label	Description	Number of tweets
1	Not beneficial	8
2	Logistical barriers	18
3	Physical harms	462
4	Not severe	3
5	Low susceptibility	1
6	Cues to action	141
7	Mistrust	90
8	Undermining religious principles	11
9	Undermining civil liberties	58
10	Additional concerns not otherwise specified	6
11	Tweet is ambiguous	18
12	No concern expressed	184
Total		1000

As can be seen in Table A1, some of the classes had fewer than 10 examples identified. Rare classes of concerns were merged based on similar themes to provide enough relevant examples to train and evaluate the performance of the multi-class classifier (Table A2).

Table A2. Number of coded tweets of each type of concern after merging the labels

Class label	Original class labels	Number of tweets
Unnecessary	1,4,5	12
Barriers	2,3	480
Cues to action	6	141
Additional concerns	7,8,9,10	165
Ambiguous	11	18
Non-concern	12	184

In the second stage, the machine learning task was a multi-class classification. A *one-versus-rest* strategy was adopted where tweets from one class (type of concern) were treated as positive samples and all other tweets were treated as negative samples. A single linear support vector machine (SVM) with stochastic gradient descent learning method was trained as the classifier for each class and this was repeated for all types of concerns. The final label for each tweet was

assigned to the class for which there was the highest likelihood of it belonging to the positive class. Given the unbalanced nature of the labeled data (some classes have a large number of tweets while several others have a small number of tweets), a random sample of 65% of the labeled tweets (the *training set*) were used to train the classifiers and the remaining 35% of the labeled tweets (the *testing/holdout set*) were used to test the performance of the classifiers. The class weights were adjusted to be inversely proportional to the number of tweets in the classes in order to mitigate the influence effect of large classes during the training. The best parameters for the classifiers were chosen using the same approach as described above.

4. Proportional exposure to HPV vaccine related tweets

The total number of followers each of the users had at the time they posted their tweets was also used to measure the potential exposure to those tweets and the potential size of the audience for each class of concerns expressed by users within each country. To quantify the potential exposure to tweets by country and type of concern, the potential exposure to each tweet was defined by the number of followers that a user had at the time they posted a tweet about HPV vaccines.

Tweets expressing concerns tended to have smaller audiences compared with tweets not expressing concern about HPV vaccines (Tables A3 and A4). In Canada, tweets expressing concerns had a total potential exposure count of 3.75% (4.81 million of 128.4 million total potential exposures to tweets from users in Canada). In Australia, the proportion was 11.0% (3.25 million of 29.7 million total potential exposures to tweets from users in Australia), and in the UK, the proportion was 16.3% (21.3 million of 130.4 million total potential exposures to tweets from users in the UK).

The difference between the number of tweets and the relative sizes of the audiences show that expressions of concern about HPV vaccines were likely to have reached a smaller overall audience than would be expected given the number of tweets. Note that these numbers reflect the total number of exposures to each type of tweet rather than the total number of unique users who may have seen those tweets.

Table A3. The number and proportion of exposures to tweets classified as expressing concerns, by country

Country	Concern	Non-concern	Total
Australia (%)	3,254,528 (10.97%)	26,422,799 (89.03%)	29,677,327 (100%)
Canada (%)	4,810,618 (3.75%)	123,608,306 (96.25%)	128,418,924 (100%)
UK (%)	21,260,539 (16.30%)	109,182,707 (83.70%)	130,443,246 (100%)
Total	29,325685 (10.16%)	259,213,812 (89.84%)	288,539,497 (100%)

Table A4. The number and proportion of exposures to tweets posted by users, by country and type of concern

Group label	Australia (%)	Canada (%)	UK (%)	Total (%)
Unnecessary	7,508 (0.03%)	9,511 (0.01%)	53,384 (0.04%)	70,403 (0.02%)

Barriers	1,950,348 (6.57%)	2,140,893 (1.67%)	9,912,306 (7.60%)	14,003,547(4.85%)
Cues to Action	287,686 (0.97%)	390,540 (0.30%)	803,537 (0.62%)	1,481,763 (0.51%)
Other concerns	595,832 (2.01%)	1,086,688 (0.85%)	1,978,772 (1.52%)	3,661,292 (1.27%)
Ambiguous	413,154 (1.39%)	1,182,986 (0.92%)	8,512,540 (6.53%)	10,108,680 (3.50%)
All non-concern	26,422,799 (89.03%)	123,608,306 (96.25%)	109,182,707 (83.70%)	259,213,812 (89.84%)
Total	29,677,327 (100%)	128,418,924(100%)	130,443,246 (100%)	288,539,497 (100%)

5. Performance of the classifiers

The binary classifier was designed to distinguish between tweets about HPV that expressed concerns from non-concerns. The binary classifier produced a precision of 90% and a recall of 90% (Table A5). In other words, approximately 1 in 10 tweets expressing a concern could have been misclassified as a non-concern tweet, and approximately 1 in 10 tweets not expressing a concern could have been misclassified as a tweet expressing a concern. Analyses reported should be interpreted in the context of this accuracy.

Table A5. Performance measures for the binary classifier within the testing/holdout set

Class label	Precision	Recall	F1 score	Number of tweets in the test set
Concern	0.90	0.97	0.93	143
Non-concern	0.89	0.74	0.81	57
Average/Total	0.90	0.90	0.90	200

The performance of the multi-class classifier varied relative to the number of instances available for training and testing in the labeled set of 1000 tweets (Table A6). The precision and recall were over 90% when identifying tweets from the 'cues to action' group, but a substantial proportion of tweets from other classes were misclassified as Class 11 (ambiguous tweets) when testing the classifier on the holdout. The performance results suggested that one could be reasonably confident about the proportions of tweets in 'barriers' and 'cues to action' groups, but less confident about the proportions of tweets belonging to other classes.

Table A6. Performance measures for the multi-class classifier within the testing/holdout set

Class label	Precision	Recall	F1 score	Number of tweets in the test set
Not beneficial, not severe, & low susceptibility	0.50	0.14	0.22	7
Barriers	0.81	0.74	0.77	180
Cues to Action	0.91	0.92	0.92	53
Other concerns	0.77	0.46	0.57	50

Ambiguous	0.03	0.4	0.06	5
Non-concern	0.45	0.31	0.37	55
Average/Total	0.74	0.64	0.68	350

6. Examination of the follower network

Examining the followers of users who expressed concerns about HPV vaccines, the results show that 34.7% of the followers of users expressing concerns were also sharing their concerns. In contrast, 8.3% of the followers of users who did not express concerns were users expressing concerns (Table A7).

Table A7. Aggregate percentages of followers for all countries and expression of concern

	Concern (%)				Non-concern (%)			
Internal network followers (aggregate follower count)	Australia	Canada	UK	All	Australia	Canada	UK	All
All concern (38,378)	4.5	17.8	12.4	34.7	9.9	18.4	37.0	65.3
All non-concern (464,251)	1.5	2.3	4.5	8.3	20.4	25.2	46.1	91.7
All users (502,629)	1.7	3.5	5.1	10.3	19.6	24.6	45.4	89.7

Examining the followers of users who expressed concerns about HPV vaccines, the results also show that these users were relatively well connected to users in other countries who also expressed concerns (Table A8). For example, 28.6% of the followers of Australian users expressing concerns were also users expressing concerns, and 52.4% of those followers were from Canada or the UK.

This type of social connection—between users from different countries—was disproportionately high between users expressing concerns about HPV vaccines, and this pattern was consistent across the three countries. These differences are also apparent in Figure 1 in the manuscript, where there is a higher density of users expressing concerns about HPV vaccines close to the boundaries between the clusters of users from Canada and the UK.

Table A8. Aggregate number and percentage of followers by country and expression of concern

	Concern (%)					Non-concern (%)				
Internal network followers (aggregate follower count)	Australia	Canada	UK	All	Proportion of international followers in the same concern group	Australia	Canada	UK	All	Proportion of international followers in the same concern group
Australian (102,894)	5.7	1.1	0.8	7.6		82.3	6.0	4.1	92.4	
-Concern (5,319)	13.6	8.3	6.7	28.6	52.4	54.4	6.5	10.5	74.4	23.8

-Non-concern (97,575)	5.3	0.7	0.4	6.4	17.2	83.8	6.0	3.7	93.6	10.4
Canadian (151,179)	0.9	9.6	1.6	12.0		6.6	71.8	9.6	88.0	
-Concern (14,656)	3.9	37.4	6.3	47.6	21.4	4.0	41.0	7.5	52.4	21.9
-Non-concern (136,523)	0.5	6.6	1.1	8.2	19.5	6.8	75.1	9.9	91.8	18.2
UK (248,556)	0.5	0.9	9.0	10.4		1.6	3.7	84.3	89.6	
-Concern (18,403)	2.3	5.0	18.8	26.1	28.0	1.8	3.9	63.1	73.9	8.3
-Non-concern (230,153)	0.4	0.5	8.2	9.1	9.9	1.6	3.7	85.6	90.9	5.8

References

1. Sloan L, Morgan J. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PloS one.* 2015;10(11):e0142209.
2. Elmongui HG, Morsy H, Mansour R. Inference models for twitter user's home location prediction. IEEE/ACS International Conference of Computer Systems and Applications (AICCSA); 2015.
3. Bo H, Paul C, Timothy B. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research.* 2014;49:451-500.
4. Bo H, Paul C, Timothy B. A. Stacking-based approach to twitter user geolocation prediction. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistic; 2013.
5. Nominatim: OpenStreetMap Wiki. http://wiki.openstreetmap.org/wiki/Nominatim.
6. Lehmann J, Gonçalves B, Ramasco JJ, Cattuto C. *Dynamical Classes of Collective Attention in Twitter.* 2012.
7. Zhao WX, Jiang J, Weng J, et al. *Comparing Twitter and Traditional Media using Topic Models.* 2011.
8. Porter MF. *An algorithm for suffix stripping.* 1997.
9. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *Journal of medical Internet research.* 2015;17(6):e144.
10. Surian D, Nguyen QD, Kennedy G, Johnson M, Coiera E, Dunn GA. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *Journal of medical Internet research.* 2016;18(8):e232.
11. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in health technology and informatics.* 2015;216:761-765.
12. Mullen T, Collier N. *Sentiment analysis using support vector machines with diverse information sources.* Barcelona Spain,2004.
13. Pang B, Lee L, Vaithyanathan S. *Thumbs up? Sentiment classification using machine learning techniques.* Philadelphia United States,2002.

14. Sebastiani F. Machine learning in automated text categorization. *ACM Comput Surv.* 2002;34:1-47.
15. Joachims T. *Transductive inference for text classification using support vector machines.* Bled Slovenia,1999.
16. McLachlan GJ, Do K, Ambroise C. *Analyzing Microarray Gene Expression Data.* Wiley; 2004.
17. Coates A, Lee H, Ng A. An analysis of single-layer networks in unsupervised feature learning. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTAT); 2011.
18. Koshiba Y, Abe S. Comparison of L1 and L2 Support Vector Machines. Proceedings of the International Join Conference on Neural Networks; 2003.

BMJ Open

Comparing human papillomavirus vaccine concerns on Twitter: A cross-sectional study of users in Australia, Canada, and the United Kingdom

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016869.R1
Article Type:	Research
Date Submitted by the Author:	30-May-2017

Complete List of Authors:	Shapiro, Gilla; McGill University, Psychology Surian, Didi; Macquarie University, Centre for Health Informatics, Australian Institute of Health Innovation Dunn, Adam; Macquarie University, Centre for Health Informatics, Australian Institute of Health Innovation Perry, Ryan; University of Melbourne, Centre for Health Policy, School of Population and Global Health Kelaher, Margaret; University of Melbourne, Centre for Health Policy, School of Population and Global Health

Primary Subject Heading:	Communication
Secondary Subject Heading:	Sexual health, Public health, Oncology, Infectious diseases, Health informatics
Keywords:	Human papillomavirus, Vaccination, Twitter, Health Belief Model, Social media

Comparing human papillomavirus vaccine concerns on Twitter: A cross-sectional study of users in Australia, Canada, and the United Kingdom

Gilla K. Shapiro (corresponding author), PhD Candidate, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada & Centre for Health Policy, School of Population and Global Health, University of Melbourne, Australia**;** Correspondence to: gilla.shapiro@mail.mcgill.ca

Didi Surian, Post-Doctoral Research Fellow, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 75 Talavera Road, Sydney, New South Wales, 2109, Australia

Adam Dunn, Senior Research Fellow**,** Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 75 Talavera Road, Sydney, New South Wales, 2109, Australia

Ryan Perry, Researcher, Centre for Health Policy, School of Population and Global Health, University of Melbourne, 235 Bouverie Street, Melbourne, Victoria, 3053, Australia

Margaret Kelaher, Professor, Head Evaluation and Implementation Science, Centre for Health Policy, School of Population and Global Health, University of Melbourne, 235 Bouverie Street, Melbourne, Victoria, 3053, Australia

Abstract

Objective: Opposition to human papillomavirus (HPV) vaccination is common on social media and has the potential to impact vaccine coverage. This study aims to conduct an international comparison of the proportions of tweets about HPV vaccines that express concerns, the types of concerns expressed, and the social connections among users posting about HPV vaccines in Australia, Canada and the United Kingdom.

Design: An international comparison of English language tweets about HPV vaccines and social connections among Twitter users posting about HPV vaccines between January 2014 and April 2016 was conducted. The Health Belief Model (HBM), one of the most widely used theories in health psychology, was used as the basis for coding the types of HPV vaccine concerns expressed on Twitter.

Setting: The content of tweets and the social connections between users who posted tweets about HPV vaccines from Australia, Canada and the United Kingdom.

Population: 16,789 Twitter users who posted 43,852 tweets about HPV vaccines.

Main outcome measures: The proportions of tweets expressing concern, the type of concern expressed, and the proportions of local and international social connections between users.

Results: Tweets expressing concerns about HPV vaccines made up 14.9% of tweets in Canada, 19.4% in Australia, and 22.6% in the UK. The types of concerns expressed were similar across the three countries, with concerns related to ‘perceived barriers’ being the most common. Users expressing concerns about HPV vaccines in each of the three countries had a relatively high proportion of international followers also expressing concerns.

Conclusions: The proportions and types of HPV vaccine concerns expressed on Twitter were similar across the three countries. Twitter users who mostly expressed concerns about HPV

vaccines were better connected to international users who shared their concerns compared to users who did not express concerns about HPV vaccines.

Key words: Human papillomavirus; Vaccination; Twitter; Health Belief Model; Social media.

Strengths and limitations of this study

- This study conducted an international comparison investigating how HPV vaccination concerns are expressed on Twitter.
- Machine learning methods were used to identify and classify the proportion and types of concerns expressed in thousands of tweets.
- The analysis of social connections among Twitter users posting about HPV vaccines in three English-speaking countries (Australia, Canada, and the United Kingdom) revealed the potential for concerns to spread internationally.
- While Twitter is used by a substantial number of people, this study is not designed to allow for generalization to the general population.
- This study used follower networks to examine social connections but further research could use other social interactions to measure the spread and impact of negative attitudes about HPV vaccines.

Introduction

Human papillomavirus (HPV) is a prevalent sexually transmitted infection that can cause cancers and anogenital warts.[1-5] Since 2006, three prophylactic vaccines have been developed to protect adolescents from HPV-associated health problems.[6] Research has demonstrated that these vaccines are safe and effective in reducing HPV related infections, genital warts, and pre-cancers.[7-12] As a result, at least sixty-five countries have implemented HPV vaccination programs for females in their national immunization schedules.[11]

There is notable variation between countries' HPV vaccine programs and coverage rates. Australia's school-based vaccination program targets 12-13 year olds females (since 2007) and males (since 2013).[13] According to Australia's National HPV vaccination program register, 85.6% of females and 77% of males received the HPV vaccine (2015 data).[14,15] In Canada, all provinces and territories introduced school-based vaccination programs for 9-13 year old females (2007-2010), and six provinces also include boys in HPV vaccine programs (since 2013).[16] According to national parental surveys, 72.3% of females (2013 data) and less than 3% of males received the HPV vaccine (2014 data).[17-19] Lastly, the United Kingdom (UK) only provides a school-based vaccination program for 12-13 year old females (since 2008). According to Public Health England, 89.5% of females in the UK received the HPV vaccine (2015 data).[20] HPV vaccine coverage rates are lower than other child or adolescent vaccines in these countries national immunisation programs;[21-23] suboptimal coverage hinders cancer prevention efforts.[24]

The media has the potential to dramatically impact vaccine coverage through influencing parental awareness, perception, and attitudes.[25-29] Unconfirmed reports of adverse events associated with the HPV vaccine published in the media dramatically affected female HPV vaccine coverage in Japan and Columbia.[11,30,31] Many individuals use the internet and social media to access health information; however, these sources have been described as a risky platform that can rapidly amplify unbalanced, distorted or inaccurate information about vaccines.[25,32-34] For example, a study by Betsch et al. found that even 5 to 10 minutes of access to vaccine-critical websites negatively influenced individuals' risk perception and intentions to be vaccinated.[35] Similarly, Nan and Madden report that, compared to a control group, participants who were exposed to negative online blogs about HPV perceived the vaccine as less safe, held more negative attitudes, and reported a reduced intention to receive the vaccine.[36]

Previous research has evaluated the public discourse concerning HPV vaccination in newspapers,[37-40] online news,[41] comments to online news articles,[42] Facebook,[43] blogs or online forums,[36,44,45] and YouTube videos.[46,47] Twitter is a microblogging service, established in 2006, that has over 313 million users active monthly. Twitter is an important source of information regarding HPV and vaccine hesitancy,[32,48,49] and several studies have examined the representation of HPV vaccines on Twitter.[33,43,50-54] Though many of these studies analyse a limited number of HPV-related tweets, a few have used data mining and machine learning techniques to analyse a large number of tweets.[51,52,54,55] However, no research has conducted an international comparison to evaluate and compare how vaccination concerns are expressed across countries. Furthermore, no research has examined the domestic and international network connectedness of HPV vaccine concern expression.

The aim of this study was to explore the proportion of HPV vaccine concern on Twitter, examine the type of concern expressed in Australia, Canada, and the United Kingdom (UK), and investigate differences in the ways Twitter users connected locally and internationally.

Methods

Study overview

Tweets related to HPV vaccines during were collected from January 2014 to April 2016 in Australia, Canada, and the UK. These countries were selected because they are English-speaking countries, share a similar history and commonwealth membership, and their similarity in administering the HPV vaccination in schools. Data captured included information about users' locations, the text of the tweets, and information about social connections. To enable the classification of a large number of tweets, two stages of machine learning classifiers were constructed from a sample of tweets that were manually coded by two investigators.

Study data

Using a similar approach to previous studies that examined large number of tweets in communities of Twitter users posting about HPV vaccines,[54,56] the Twitter Search Application Programming Interface (API) was used to collect tweets in the English language about HPV vaccines from January 2014 to April, 2016. The search terms were "Gardasil", "Cervarix", "hpv AND vaccin*", and "cervical AND vaccin*". Information extracted from each tweet included the unique tweet identifier, tweet text, creation time, the identifier of the user posting the tweet, and geographical coordinates (if available). Without any restrictions applied to the locations of users, the entire dataset included 358,194 tweets (including retweets) by 129,286 users.

A gazetteer was used to transform the text provided by users into coordinates, and any users with self-reported locations that were located in coordinates in Australia, Canada, or the UK were included in this analysis (Supplemental Material, Section 1).

Other data that were used in the analyses included the set of social connections formed among the users who were included in the analyses. For each user, the Twitter Search API was used to collect the set of all follower relationships in which the user was involved, shortly after the first time the user posted a relevant tweet in the period. A network was then formed to include all users who tweeted about HPV vaccines from the three countries, and the follower relationships defined the social connections in an unweighted, directed network (Supplemental Material, Section 2).

The Macquarie University Human Research Ethics Committee (#5201401028) and the University of Melbourne's Research Ethics Board (#1647488.1) provided ethics approval for data collection and analysis.

Analysis

Supervised machine learning methods were used to classify the tweets in two stages; firstly to identify tweets that expressed any concern and secondly to classify specific types of concerns. In the first stage, 1000 tweets were sampled from the set of all tweets to manually label those that expressed concerns. In the second stage, 1000 tweets were sampled from the set of tweets that were estimated to be concerns to manually label them by type. The manually labelled tweets were used to train classifiers to label any tweet by the type of concern expressed (Supplemental Material, Section 3).

The categories for types of tweets expressing concerns were determined using an inductive and deductive procedure. The Health Belief Model (HBM), one of the most widely used theories in health psychology, was used as the basis for coding the types of HPV vaccine concerns expressed on Twitter.[57] The HBM has been used previously to evaluate the determinants of HPV vaccination and non-compliance by identifying perceived susceptibility to HPV, perceived severity of HPV, perceived benefits of HPV vaccination, perceived barriers of HPV vaccination (including tangible barriers such as logistical challenges and psychological barriers such as perceived harms of receiving the HPV vaccine), and cues to action (e.g. influences prompting HPV vaccine uptake such as information from health care provides, family, or friends).[47,57-59] To account for additional prominent concerns that were not captured by the model, the constructs were also informed by previous content analyses of media and social media related to the HPV vaccine, as well as literature on vaccine hesitancy.[28,33,37,41,42,45,47,60-68] The coding scheme was therefore extended to include mistrust, undermining of religious principles, undermining of civil liberties, additional concerns (not otherwise specified), and ambiguous tweets (Table 1). The coding scheme was used by two investigators (GS and RP) to code 12 types of concerns expressed in a second sample of 1000 tweets. After examining the proportions of different types of concerns in the sample and accuracy of the multi-class classifier (Supplemental Material, Sections 3 and 4), types of concerns were combined to improve the performance that could be achieved by the machine learning classifiers (Table 1). Combining categories was done based on conceptual similarity and trying to remain as true to the HBM as possible while attaining accuracy of the classifier.

The network of social connections formed by users who posted tweets about HPV vaccines was used to compare the proportions of local (within a country) and international (across countries) followers. The group of users for whom at least half of their relevant tweets were expressing concerns were assigned to one group (concern), and all other users were assigned to another (non-concern). The users were then also split by country, and the proportions of local and international followers were compared across groups.

Results

There were 129,286 Twitter users who posted at least one tweet about HPV vaccines during the period. The location inference method identified 2,792 (2.2%) of those users located in Australia, 7,237 (5.6%) located in Canada, and 6,760 (5.2%) users located in the UK (Table 2).

From the 16,789 users in the three countries, a total of 43,852 tweets about HPV vaccines were posted, of which 7,173 (16.4%) were from Australia, 18,927 (43.2%) were from Canada, and 17,752 (40.5%) were from the UK. This corresponded to an average of 2.57 tweets per user in Australia (range, 1-198), 2.61 tweets per user in Canada (range, 1-433), and 2.62 tweets per user in the UK (range, 1-501).

Expressions of concern

When labelling tweets that expressed concerns, the binary classifier (stage one) achieved a recall of 0.97 and a precision of 0.90. This indicates that the binary classifier missed 3% of tweets that were manually labelled as having expressed a concern and 10% of tweets it labelled as having expressed a concern were manually labelled otherwise. Because the multi-class classifier (at stage two) identified a proportion of these mislabelled tweets in the second round, the overall rate of error in stage one was within 5% of the correct proportion (Supplemental Material, Section 4).

The proportion of tweets posted about HPV vaccines from users in the three countries expressing concerns was 18.7% (8,215 of 43,852 tweets), but there were differences in these proportions across the three countries (Table 2). Canada had the lowest proportion of tweets expressing concerns at 14.9% (2,818 of 18,927 tweets), followed by Australia at 19.3% (1,388 of 7,173 tweets). The UK had the highest proportion of tweets expressing concerns at 22.6% (4,009 of 17,752 tweets). Tweets expressing concerns also tended to have smaller audiences compared with tweets not expressing concern about HPV vaccines (Supplemental Material, Section 3).

Types of concerns expressed

When identifying concerns related to cues to action the classifier respectively produced a precision of 0.81 and a recall of 0.74. For perceived barriers, the precision was 0.91 and recall was 0.92. The classifier was less reliable for the remainder of the concern groups because these types of concerns made up a much smaller proportions, resulting in imbalance in the data, which affects the performance that can be achieved by the classifiers (Supplemental Material, Section 5).

Tweets expressing concerns about perceived barriers comprised the largest type of concern by both the proportion of tweets expressing concerns (Table 3). The proportions of each group of concerns across the three countries were generally consistent.

Social connections among users

Among users from the three countries who posted about HPV vaccines, 18.2% (3,062 of 16,789) were labelled as having expressed concerns (at least half of the tweets about HPV vaccines they posted were labelled as having expressed a concern). The total number of follower connections among the set of 16,789 users was 502,629. Users from the three countries were disproportionately more likely to be followed by users from the same country, creating clusters of users by country (Figure 1). Furthermore, users who expressed concerns about the HPV vaccines appear to be more tightly connected within the United Kingdom, compared to either Australia or Canada. Figure 1 also highlights that users discussing HPV vaccines in the United Kingdom are more often connected to users in Australia and Canada than users in Australia and Canada are connected to each other.

To examine the proportion of followers of HPV vaccine tweets, Figure 2 examines “concern” and “non-concern” tweets for each of the three countries (to produce six groups represented as circles). Relative to users who did not express concern about HPV, users that did express concerns had a higher proportion of international followers who also expressed concerns

(Figure 2). Among UK users expressing concerns, 26.1% of followers also expressed concerns (compared to 9.1% of followers among UK users not expressing concerns). Also among UK users expressing concerns, 28.0% of their followers also expressed concerns and were from Australia or Canada, and 9.9% of their followers did not express concerns and were from Australia or Canada. In comparison, among UK users not expressing concerns, only 5.8% of their followers were users not expressing concerns and from Australia or Canada, and 8.3% of their followers were users expressing concerns and from Australia or Canada (Supplemental Material, Section 6). This pattern was consistent across each of the three countries. The results indicate that users who mostly expressed concerns were disproportionately well-connected to international users discussing HPV vaccines.

Discussion

This study found that in Australia, Canada, and the UK nearly one in five of the tweets about HPV vaccines were expressions of concern. Canadian Twitter users less often expressed concerns about HPV vaccines (14.9%) compared to Australia (19.3%) and the UK (22.6%) (Table 2). There was a general consistency in the proportions of specific concerns across the three countries, and the most common concerns (46%) were related to 'perceived barriers' (i.e. logistical challenges and psychological barriers such as perceived harms of receiving the HPV vaccine) (Table 3). The results demonstrated that users expressing concerns about HPV vaccines tended to be relatively well-connected to users discussing HPV vaccine concerns in other countries, especially between Canada and the UK.

Previous studies examining the representation of HPV vaccines on Twitter identified slightly higher proportions of negative tweets or tweets expressing concerns, but these studies captured different time periods and did not compare specific countries.[52,56] For example, a study of six months of Twitter data in the United States (between October 2013 and April

2014) found 25.1% of tweets were negative.[51] Though greater research is required, the balance of positive and negative content appears to vary by source whereby the majority of news content,[37] online comments (in response to news articles),[42] and tweets have been found to be positive; the majority of YouTube content has been found to be negative.[47] In examining the type of concern expressed about the HPV vaccine on other social media sites, researchers have also observed the predominance of perceived barrier (i.e. logistical challenges and psychological barriers such as perceived harms of receiving the HPV vaccine).[44,45,69] However, while the present research study found concerns about safety were most common on Twitter; other research found safety to be surpassed or similar in salience to other prevalent themes including conspiracies/search for truth, mistrust for health system, and promoting promiscuity.[44,45,69] Surian et al. analysed topics regarding HPV vaccines on Twitter and found individuals who posted about 'harms and conspiracies' posted more often than other users, suggesting that some users are actively seeking to introduce concerns about HPV vaccines into the public domain.[54] The predominance of HPV vaccine concerns about perceived barrier on Twitter indicates the importance of these concerns. It would be valuable to extend this work to examine differences in general vaccine concerns as well as compare concerns towards specific vaccines on Twitter.

International networking on Twitter suggests that vaccine related controversies in one country could reverberate around the world and impact vaccine coverage. Public health professionals and policymakers must therefore be able to monitor, rapidly identify, and react to such concerns (e.g. by providing evidence-based responses in real-time and strengthening their own international networks).[34,50,70] This research provides public health practitioners and policymakers with evidence that concerns about 'perceived barriers' on Twitter are widespread; effective communication campaigns could be designed and implemented to target this concern in locations where it is likely to have the greatest impact. However, it is important for further research to analyse results by type of sender. It would also be critical for future research to design and evaluate appropriate messaging of such a campaign so that this intervention does not 'backfire' and increase hesitancy.[71]

This study also found that Twitter users expressing HPV vaccine concerns tended to have higher proportions of international connections compared to those not expressing concerns. Public health organizations seeking to improve the uptake of HPV vaccines may benefit from tools that help them monitor the impact of vaccine scares on social media in other countries in order to pre-empt and respond to misinformation locally. Such support could have been beneficial for Japan and Colombia when the media had a detrimental impact on HPV vaccine coverage.[11,30,31]

Similar to other studies in this area, our research did not measure whether the expression of concerns on Twitter led to changes in decision-making and coverage. Further research would be beneficial to assess the pathway of HPV vaccine concerns, and whether such concerns have a real-world impact (e.g. on vaccine coverage). In particular, future work should consider the relationship between the information about HPV vaccines that enters into the public discourse and the decision-making of individuals and populations. While Canada had the lowest proportion of tweets in which concerns were expressed of the three countries, it also has the lowest rate of HPV vaccine uptake.[14,15,18,19,72] Accordingly, we echo Gollust et al.'s recent call for greater experimental research designs to make causal assertions about the impact of the media on vaccine coverage.[73] Although some studies have begun to do so,[36,74] it would be helpful for future research to specifically evaluate the impact of Twitter messages and for moderating variables to also be evaluated.

Together with other studies on the representation of HPV vaccines in the media, our results suggest that it would be useful to monitor early indications of negative influence on attitudes and beliefs on social media. Two studies have independently examined responses on Twitter to specific controversial events including US Representative Michele Bachmann's claim that HPV vaccines could cause "mental retardation", and Katie Couric's television segment "HPV Vaccine Controversy" that aired on December 4, 2013 [53,75]. Mahoney *et al.* (2015) evaluated 200 social media posts before and after Bachmann's comments on the *Today Show* and found that though most media was positive in tone, compared to Google News, Twitter disseminated more positive HPV vaccine articles and also used more personal accounts as a reference source [53]. In contrast, using a random sample of 3,595 tweets, Bahk *et al.* (2016) found that most sentiment on Twitter towards HPV vaccines before Katie Couric's episode was negative, and while there was a decrease of negative sentiment immediately after the show aired, negative sentiment returned to baseline after two weeks [75]. Future research should also investigate how public health organizations should effectively intervene to curb misinformation or 'fake news' regarding HPV vaccination.

There were several limitations to this study. First, the findings are specific to Twitter, and while Twitter represents one of the largest populations of social media users, the results are not necessarily representative of the broader public discourse about HPV vaccines in news and online social media.[76-79] Twitter is an inherently biased representation of the broader population, and is skewed both in terms of age and socioeconomics.[80-83] Second, while the location inference method is a standard in the area,[84,85] the methods are imperfect.[86-89] Third, the study was limited to English-language tweets in three countries and evaluations of other countries and other languages may have yielded different results. It would be beneficial for future research to expand the focus of analysis to examine diverse countries, as well as conduct more nuanced regional explorations of a single country. Finally, using networks based on which users follow each other does not necessarily capture all of the interactions that occur online. While some argue that interaction with content (liking or retweeting) is a better measure of impact than followers,[90] others have argued that many users on Twitter are passive and do not interact with the content,[91] and as such followers may be a better indicator of impact. As this study examined follower networks, it would be helpful for future research to compare followers to different ways of interacting with content in order to better understand the impact of HPV vaccine tweets.

Conclusions

This study characterized the concerns about HPV vaccines expressed by Twitter users in three countries. The UK had the greatest proportion of tweets expressing concerns about HPV vaccines and Canada had the least, and the types of concerns expressed were relatively consistent across the three countries. Users who expressed concerns about HPV vaccines were generally more closely connected to users in other countries who also expressed concerns, suggesting that controversies and misinformation may be rapidly shared across international boundaries. This research could be used to design public health interventions that address concerns about the HPV vaccine on Twitter. In particular, this study suggests that methods for addressing vaccine concerns may benefit from targeting concerns about perceived barriers to vaccination (including logistical challenges and psychological barriers such as vaccine pain, safety, and side effects as a consequence of receiving the HPV vaccine). In addition, further coordination of public health agencies internationally may mitigate vaccine scares.

Acknowledgement

Contributors GS, AD, and MK conceptualized and designed the study. DS and AD acquired the data. GS drafted the manuscript. All authors analysed and interpreted the data, critically revised the manuscript for important intellectual content, and approved the final version.

Funding: GS was supported by the CIHR Vanier Canada Graduate Scholarship and Queen Elizabeth II Diamond Jubilee Scholarship programs. AD is an investigator on a related NHMRC Project Grant (APP1128968). The funding source had no role in the design of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

Competing interests: None.

Ethics approval: Macquarie University Human Research Ethics Committee (#5201401028) and the University of Melbourne's Research Ethics Board (#1647488.1).

Data sharing statement: No additional data is available. Due to ethical approvals and the Twitter Terms of Service, the individual identifiers for Twitter posts and users cannot be provided with their manual or automatic labels.

References

1. Trottier H, Franco EL. The epidemiology of genital human papillomavirus infection. *Vaccine.* 2006;24(1):S1–15.
2. Tota JE, Chevarie-Davis M, Richardson LA, Devries M, Franco EL. Epidemiology and burden of HPV infection and related diseases: implications for prevention strategies. *Prev Med* 2011;53(1):S12-21.
3. Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *The Lancet Global Health.* 2016;4(9):e609–e616
4. Forman D, de Martel C, Lacey CJ, et al. Global burden of human papillomavirus and related diseases. *Vaccine.* Nov 20 2012;30 Suppl 5:F12-23.
5. National Advisory Committee on Immunization. Update On Human Papillomavirus Vaccines. *Canada Communicable Disease Report.* 2012;38.
6. Canadian Cancer Society. *Canadian Cancer Statistics 2016.* 2016.
7. Markowitz LE, Dunne EF, Saraiya M, et al. Human papillomavirus vaccination: recommendations of the Advisory Committee on Immunization Practices. *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control.* Aug 29 2014;63(Rr-05):1-30.
8. Ferris D, Samakoses R, Block SL, et al. Long-term study of a quadrivalent human papillomavirus vaccine. *Pediatrics.* Sep 2014;134(3):e657-665.
9. Garland SM. The Australian experience with the human papillomavirus vaccine. *Clin Ther* 2014;36:17-23.

10. Brotherton JM, Saville AM, May CL, Chappell G, Gertig DM. Human papillomavirus vaccination is changing the epidemiology of high-grade cervical lesions in Australia. *Cancer Causes Control.* 2015;26(6):953–954.

11. Brotherton JML, Bloem PJN. HPV Vaccination: Current Global Status. *Curr Obstet Gynecol Rep.* 2015;4:220-233.

12. Drolet M, Benard E, Boily MC, et al. Population-level impact and herd effects following human papillomavirus vaccination programmes: A systematic review and meta-analysis *Lancet Infect Dis.* 2015;15(5):565-580.

13. National Health Performance Authority. *Healthy Communities: HPV immunisation rates for girls in 2013 (In Focus)* 2015.

14. National HPV Vaccination Program Register. *National (Australia) HPV 3 dose vaccination coverage for females turning 15 years of age in 2015.* 2016.

15. National HPV Vaccination Program Register. National (Australia) HPV 3 dose vaccination coverage for males turning 15 years of age in 2015. 2016.

16. Shapiro Gilla K, Perez Samara, Rosberger Zeev. Including males in Canadian human papillomavirus vaccination programs: a policy analysis. *CMAJ.* 2016;188(12):881-886.

17. Samara Perez, Gilla K. Shapiro, Christopher A. Brown, Eve Dube, Gina Ogilvie, Rosberger Z. 'I didn't even know boys could get the vaccine': Parents' reasons for human papillomavirus (HPV) vaccination decision making for their sons. *Psycho-Oncology.* 2015.

18. Perez S, Tatar O, Shapiro GK, et al. Psychosocial determinants of parental human papillomavirus (HPV) vaccine decision-making for sons: Methodological challenges and initial results of a pan-Canadian longitudinal study. *BMC Public Health.* 2016;16(1):1223.

19. Gilbert NL, Gilmour H, Dubé E, Wilson SE, Laroche J. Estimates and determinants of HPV non-vaccination and vaccine refusal in girls 12 to 14 years of age in Canada: Results from the Childhood National Immunization Coverage Survey, 2013. *Human Vaccines & Immunotherapeutics.* 2016;12(6):1484-1490.

20. Bhavita Vishram, Lisa Byrne, Joanne White, Michael Edelstein. *Human Papillomavirus (HPV) vaccination coverage in adolescent females in England: 2014/15.* London, England: Public Health England;2015.

21. Australian Government Department of Health. Australian Immunisation Register: Current Data. In: Program IA, ed: Commonwealth of Australia; 2017.

22. Screening & Immunisations Team-NHS Digital. *NHS Immunisation Statistics England, 2015-16.* 2016.

23. Government of Canada. Vaccine coverage in Canadian children: Highlights from the 2013 childhood National Immunization Coverage Survey (cNICS). 2016.

24. Holman DM, Benard V, Roland KB, Watson M, Liddon N, Stokley S. Barriers to Human Papillomavirus Vaccination Among US Adolescents: A Systematic Review of the Literature. *JAMA pediatrics.* 2014; 168 (1):76-82.

25. Kumar D, Chandra R, Mathur M, Samdariya S, Kapoor N. Vaccine hesitancy: Understanding better to address better. *Isr J Health Policy Res.* 2016;5(2).

26. Mason BW, Donnelly PD. Impact of a local newspaper campaign on the uptake of the measles mumps and rubella vaccine. *J Epidemiol Community Health.* 2000;54:473-474.

27. Gangarosa EJ, Galazka AM, Wolfe CR, et al. Impact of anti-vaccine movements on pertussis control: the untold story. *Lancet.* 1998;351(9099):356-361.

28. Larson HJ, Jarrett C, Schulz WS, et al. Measuring vaccine hesitancy: The development of a survey tool. *Vaccine.* 2015.

29. Gollust SE, Attanasio L, Dempsey A, Benson AM, Fowler EF. Political and news media factors shaping public awareness of the HPV vaccine. *Womens Health Issues.* 2013;23(3):e143-151.

30. Hanley SJ, Yoshioka E, Ito Y, Kishi R. HPV vaccination crisis in Japan. *Lancet.* 2015;385(9987):2571.

31. El Tiempo. Caso de niñas del Carmen de Bolívar desplomó vacunación contra el VPH. February 14, 2015; http://www.eltiempo.com/estilo-de-vida/salud/vacuna-contra-el-papiloma-humano-cayo-58-puntos-porcentuales/15246061.

32. Gabarron E, Serrano JA, Wynn R, Lau AYS. Tweet content related to sexually transmitted diseases: no joking matter. *Journal of medical Internet research.* 2014;16(10):e228.

33. Britt RK, Hatten KN, Chappuis SO. Perceived behavioral control, intention to get vaccinated, and usage of online information about the human papillomavirus vaccine. *Health psychology and behavioral medicine.* 2014;2(1):52-65.

34. Amicizia D, Domnich A, Gasparini R, Bragazzi NL, Lai PL, Panatto D. An overview of current and potential use of information and communication technologies for immunization promotion among adolescents. *Human Vaccines & Immunotherapeutics* 2013;9(12):2634–2642.

35. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. *J Health Psychol.* 2010;15:446–455.

36. Nan X, Madden K. HPV vaccine information in the blogosphere: how positive and negative blogs influence vaccine-related risk perceptions, attitudes, and behavioral intentions. *Health communication.* 2012;27(8):829-836.

37. Perez S, Fedouruk C, Shapiro GK, Rosberger Z. Giving Boys a Shot: The HPV Vaccine's Portrayal in Canadian Newspapers. *Health communication.* 2016;31(12):1527-1538.

38. Robbins SC, Pang C, Leask J. Australian newspaper coverage of human papillomavirus vaccination, October 2006-December 2009. *Journal of health communication.* 2012;17(2):149-159.

39. Krieger JL, Katz ML, Eisenberg D, Heaner S, Sarge M, Jain P. Media coverage of cervical cancer and the HPV vaccine: implications for geographic health inequities. *Health expectations : an international journal of public participation in health care and health policy.* Sep 2013;16(3):e1-12.

40. Quintero JJ, Sionean C, Scott AM. Exploring the presentation of news information about the HPV vaccine: a content analysis of a representative sample of U.S. newspaper articles. *Health communication.* 2011;26(6):491-501.

41. Habel MA, Liddon N, Stryker JE. The HPV vaccine: A content analysis of online news stories. *J Womens Health.* 2009;18(3):401-407.

42. Feinberg Y, Pereira JA, Quach S, et al. Understanding Public Perceptions of the HPV Vaccination Based on Online Comments to Canadian News Articles. *PloS one.* 2015;10(6):e0129587.

43. Coloma PM, Becker B, Sturkenboom MC, van Mulligen EM, Kors JA. Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies. *Drug safety.* Oct 2015;38(10):921-930.

44. Penta MA, Baban A. Dangerous agent or saviour? HPV vaccine representations on online discussion forums in Romania. *International journal of behavioral medicine.* 2014;21(1):20-28.

45. Keelan J, Pavri V, Balakrishnan R, Wilson K. An analysis of the Human Papilloma Virus vaccine debate on MySpace blogs. *Vaccine.* Feb 10 2010;28(6):1535-1540.

46. Ache KA, Wallace LS. Human Papillomavirus vaccination coverage on YouTube. *Am J Prev Med* 2008;35(4):389-392.

47. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. *Health communication.* 2012;27:478–485.

48. Gunasekaran B, Jayasinghe Y, Fenner Y, et al. Knowledge of human papillomavirus and cervical cancer among young women recruited using a social networking site. *Sexually transmitted infections.* 2013;89(4):327-329.

49. McNeill A, Harris PR, Briggs P. Twitter Influence on UK Vaccination and Antiviral Uptake during the 2009 H1N1 Pandemic. *Frontiers in public health.* 2016;4:26.

50. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly available online tool facilitates real-time monitoring of vaccine conversations and sentiments. *Health Affairs.* 2016;35(2):341-347.

51. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *Journal of medical Internet research.* 2015;17(6):e144.

52. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in health technology and informatics.* 2015;216:761-765.

53. Mahoney LM, Tang T, Ji K, Ulrich-Schad J. The Digital Distribution of Public Health News Surrounding the Human Papillomavirus Vaccination: A Longitudinal Infodemiology Study. *JMIR Public Health Surveill.* 2015;1(39):e2.

54. Surian D, Nguyen QD, Kennedy G, Johnson M, Coiera E, Dunn GA. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *Journal of medical Internet research.* 08/29 2016;18(8):e232.

55. Salathé M, Vu DQ, Khandelwal S, Hunter DR. The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science.* 2013.

56. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *Journal of Medical Internet Research.* 2015;17(6):e144.

57. Rosenstock IM. Historical origins of the health belief model *Health Educ Monogr* 1974;2:328-335.

58. Donadiki EM, Jiménez-García R, Hernández-Barrera V, et al. Health Belief Model applied to non-compliance with HPV vaccine among female university students. *Public Health.* 2014;128(3):268-273.

59. Krawczyk A, Knäuper B, Gilca V, et al. Parents' decision-making about the human papillomavirus vaccine for their daughters: I. Quantitative results. *Hum Vaccin Immunother.* 2015;11(2):322-329.

60. Larson H, Leask J, Aggett S, Sevdalis N, Thomson A. A Multidisciplinary Research Agenda for Understanding Vaccine-Related Decisions. *Vaccines.* 2013;1(3):293-304.

61. Casciotti DM, Smith KC, Tsui A, Klassen AC. Discussions of adolescent sexuality in news media coverage of the HPV vaccine. *Journal of adolescence.* 2014;37(2):133-143.

62. Casciotti DM, Smith KC, Andon L, Vernick J, Tsui A, Klassen AC. Print news coverage of school-based human papillomavirus vaccine mandates. *The Journal of school health.* 2014;84(2):71-81.

63. Madden K, Nan X, Briones R, Waks L. Sorting through search results: a content analysis of HPV vaccine information online. *Vaccine.* 2012;30(25):3741-3746.

64. Nan X, Daily K. Biased assimilation and need for closure: Examining the effects of mixed blogs on vaccine-related beliefs. *Journal of health communication.* 2015;20(4):462-471.

65. Thomson A, Robinson K, Vallee-Tourangeau G. The 5As: A practical taxonomy for the determinants of vaccine uptake. *Vaccine.* 2016;34(8):1018-1024.

66. Wolfe RM, Sharp LK, Lipsky MS. Content and design attributes of antivaccination web sites. *JAMA.* 2002;287(24):3245-3248.

67. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: A content analysis. *JAMA.* 2007;298(21):2482-2484.

68. Abdelmutti N, Hoffman-Goetz L. Risk messages about HPV, cervical cancer, and the HPV vaccine Gardasil: A content analysis of Canadian and U.S. national newspaper articles. *Women & health.* Jul-Aug 2009;49(5):422-440.

69. Kata A. A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine.* 2010;28:1709-1716.

70. Larson HJ, Wilson R, Hanley S, Parys A, Paterson P. Tracking the global spread of vaccine sentiments: The global response to Japan's suspension of its HPV vaccine recommendation. *Hum Vaccin Immunother.* 2014;10(9):2543-2550.

71. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: A randomized trial. *Pediatrics.* 2014;133:e835-e842.

72. Vishram B, Byrne L, White J, Edelstein M. *Human Papillomavirus (HPV) vaccination coverage in adolescent females in England: 2014/15.* London, England: Public Health England;2015.

73. Gollust SE, LoRusso SM, Nagler RH, Fowlerc EF. Understanding the role of the news media in HPV vaccine uptake in the United States: Synthesis and commentary. *Hum Vaccin Immunother.* 2016;12(6):1430-1434.

74. Betsch C. Social media targeting of health messages: A promising approach for research and practice. *Hum Vaccin Immunother.* 2014;10(9):2636-2637.

75. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly Available Online Tool Facilitates Real-Time Monitoring Of Vaccine Conversations And Sentiments. *Health affairs (Project Hope).* Feb 2016;35(2):341-347.

76. Weeg C, Schwartz HA, Hill S, Merchant RM, Arango C, Ungar L. Using Twitter to Measure Public Discussion of Diseases: A Case Study. *JMIR Public Health Surveill.* 2015;1(1):e6.

77. Pew Research Center. *Social networking fact sheet.* http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/.

78. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in twitter: A quantitative analysis. *JMIR Public Health Surveill.* 2016;2(1):e1.

79. Paul MJ, Dredze M. You are what you tweet: Analyzing twitter for public health. *Association for the Advancement of Artificial Intelligence* 2011.

80. Sadah AS, Shahbazi M, Wiley TM ea. A study of the demographics of web-based health-related social media users. *Journal of medical Internet research.* 2015;17(8):e194.

81. Mislove A, Lehmann S, Ahn Y-Y, Onnela J-P, Rosenquist J.N. Understanding the demographics of Twitter users. Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM); 2011.

82. Sloan L, Morgan J, Burnap P, Williams M. Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data. . *PloS one*. 2015;10(3):e0115545.

83. Sloan L, Morgan J. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PloS one.* 2015;10(11):e0142209.

84. Kumar S, Morstatter F, Marshall G, Liu H, Nambiar U. Navigating Information Facets on Twitter (NIF-T). Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD); 2012.

85. Kounadi O, Lampoltshammer TJ, Groff E, Sitko I, Leitner M. Exploring twitter to analyze the public's reaction patterns to recently reported homicides in London. *PloS one.* 2015;10(3):e0121848.

86. Elmongui HG, Morsy H, Mansour R. Inference models for twitter user's home location prediction. IEEE/ACS International Conference of Computer Systems and Applications (AICCSA); 2015.

87. Bo H, Paul C, Timothy B. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research.* 2014;49:451-500.

88. Jurgens D. That's what friends are for: Inferring location in online social media platforms based on social relationships. . Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM); 2013.

89. Cheng Z, Caverlee J, Lee K. You are where you tweet: A content-based approach to geo-locating Twitter users. Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM); 2010.

90. Cha M, Haddadi H, Benevenuto F, Gummadi KP. Measuring user influence in Twitter: The million follower fallacy. Association for the Advancement of Artificial Intelligence 2010.

91. Romero DM, Galuba W, Asur S, Huberman BA. Influence and passivity in social media. In: Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, Michalis Vazirgiannis, eds. *Machine Learning and Knowledge Discovery in Databases* 2011:18-33.

Tables and Figures

Table 1. Coding scheme for the types of concerns expressed on Twitter

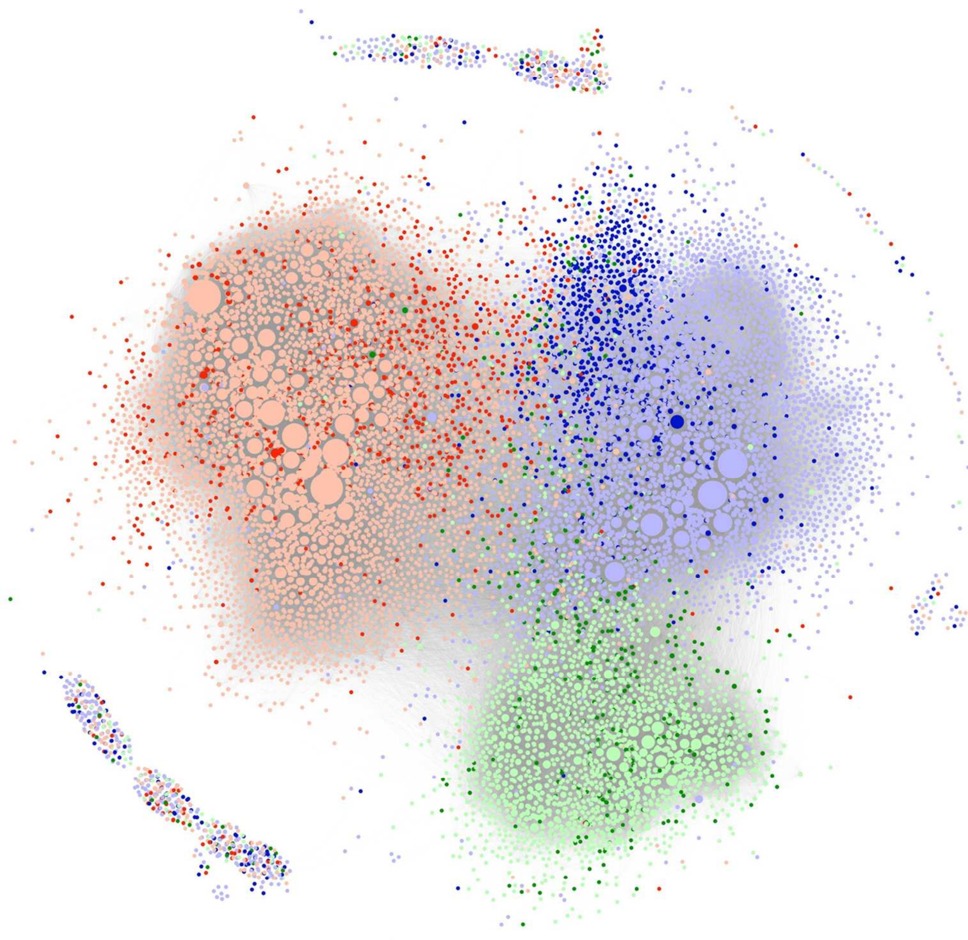
Original 12 types of concern	Combined concern group label	Example tweet
Not beneficial: stating that the HPV vaccine is not beneficial or useful (HBM construct 'perceived benefits')	Unnecessary	"Breaking Report HPV Cancers Rising In Spite of Vaccination URL #vaccination #guardasil #cervarix #HPV #cancer #fraud"
Perceived logistical challenges: stating logistical barriers such accessible or affordability challenges (HBM construct 'perceived barriers')	Perceived barriers	"Makes no sense that girls are covered for the HPV vaccine but I gotta pay $400 for it... #needstochange"
Perceived harms: stating concerns about the physical issues or harms as a result of receiving the HPV vaccine including pain, safety, or side effects (HBM construct 'perceived barriers')	Perceived barriers	"Mum still reckons im unwell like this all the time cos of the hpv vaccine. Thinking she might be right. Hpv vacc has loads of side effects"
Not severe: stating that HPV and/or its consequences are not severe (e.g. because it is common or clears up on its own) (HBM construct 'perceived severity')	Unnecessary	"The New Gardasil Is It Right For Your Daughter URL"
Low susceptibility: stating the HPV vaccine is unnecessary because there is a low likelihood of getting HPV and/or its consequences (HBM construct 'perceived susceptibility')	Unnecessary	"30 Facts you probably don\u2019t know about HPV and Gardasil...URL"
Cues to action: stating the influence of significant others guiding against receiving HPV vaccination (HBM construct 'cues to action')	Cues to action	"American College of Pediatricians warns about toxic effects of Gardasil vaccine"
Mistrust: stating a lack of confidence, mistrust, scepticism or belief in a HPV vaccine conspiracy	Additional concern	"Save dosh on the pharmaceutically lucrative, dubious Gardasil vaccine #qanda"
Undermining religious principles: stating concern that the HPV vaccine is inconsistent with religious principles	Additional concern	"…b. c. bishop, says chastity, not hpv vaccine, will keep girls healthy…"
Undermining civil liberties: stating concern about civil liberties (e.g. girls-only mandate, autonomy, who should be the decision maker for child vaccination, not being adequately consulted etc.)	Additional concern	"…had one dose of the gardasil at 17 after being bullied into it by my doctor, he basically told me I wasn't leaving without it"
Additional concerns not otherwise specified (e.g. belief in alternative medicine)	Additional concern	"…Our body does not need something NOT natural in our body to heal! The Gardasil/Vaccines were all in the..."
Tweet is ambiguous	Ambiguous	"…oh well if it's peer reviewed I'll give my son a gardasil shot."
No concern expressed	Non-concern	"Just saw a commercial that was like ask your doctor about Gardasil and I pumped my fist and shouted already did! because #sexualhealth"

Table 2. The total number of users and tweets from Australia, Canada, and the UK

Country	Number of users	Number of tweets	Tweets expressing concern	Tweets not expressing concern
Australia (%)	2,792 (16.6%)	7,173 (16.4%)	1,388 (19.4%)	5,785 (80.6%)
Canada (%)	7,237 (43.1%)	18,927 (43.2%)	2,818 (14.9%)	16,109 (85.1%)
UK (%)	6,760 (40.3%)	17,752 (40.5%)	4,009 (22.6%)	13,743 (77.4%)
Total	16,789 (100%)	43,852 (100%)	8,215 (18.7%)	35,637 (81.3%)

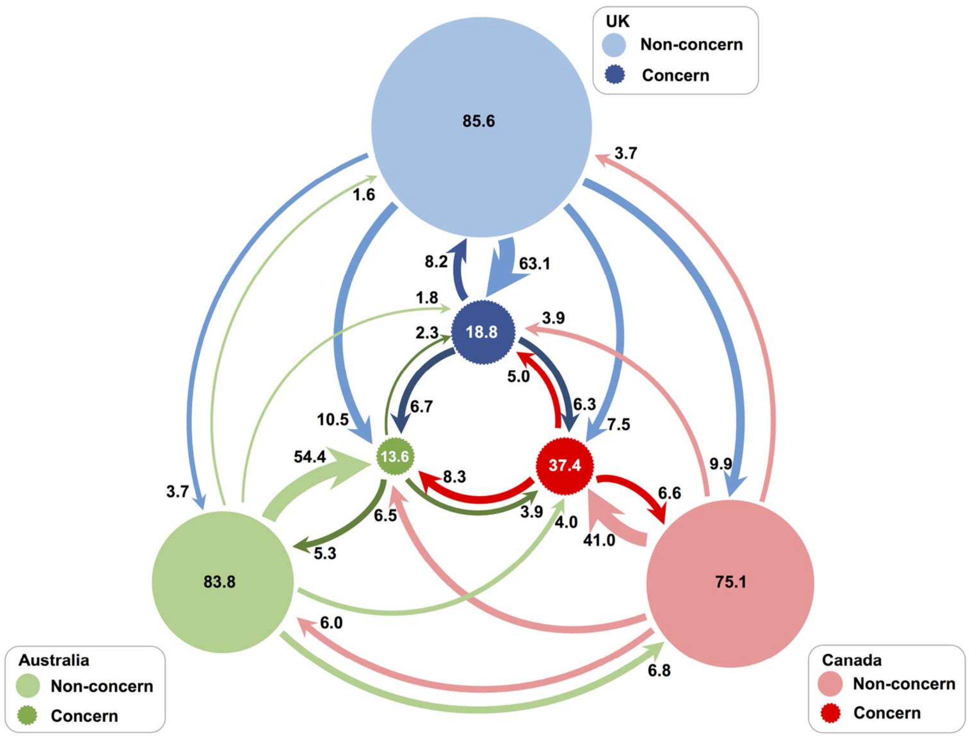
Table 3. Number of tweets, by country and concern type

Group label	Australia (%)	Canada (%)	UK (%)	Total (%)
Unnecessary	6 (0.39%)	13 (0.4%)	29 (0.6%)	48 (0.5%)
Perceived barriers	717 (47.08%)	1,368 (42.4%)	2,137 (48.0%)	4,222 (45.9%)
Cues to Action	157 (10.31%)	274 (8.5%)	306 (6.9%)	737 (8.0%)
Additional concerns	187 (12.28%)	469 (14.5%)	560 (12.6%)	1,216 (13.2%)
Ambiguous	321 (21.08%)	694 (21.5%)	977 (22.0%)	1,992 (21.7%)
Total concern	1,388 (100%)	2,818 (100 %)	4,009 (100%)	8,215 (100%)



The follower network for Twitter users posting about HPV vaccines is coloured by country (Australia, green; Canada, red; UK, blue). Each node represents a user, and the node sizes are proportional to the number of followers within the user's network. Nodes are positioned by heuristic to be closer to nodes with which they are better connected, as a way of illustrating the community structure. Darker coloured nodes indicate users for whom at least 50% of their relevant tweets expressed concerns.

127x127mm (300 x 300 DPI)



The percentages of followers for all users by expression of concern from Australia, Canada and the UK. The circle represents a concern group of Twitter users, where the circle size is proportional to the number of users. The arrow represents user following direction. The number represents the percentage of followers, where the number in a circle represents the percentage of followers from the same concern group. Only values above 1.5% are shown, see Supplemental Material (Section 5) for all values.

146x112mm (300 x 300 DPI)

Supplemental Material

The information described below provides greater detail of the methodology and results described in the manuscript.

1. Extracting and estimating the home locations of Twitter users

The use of 'geo-tags' on Twitter—where geographical coordinates are embedded in the metadata—is relatively rare and unevenly distributed, which makes geo-tags an unreliable way to determine the locations of users.[1] A common alternative is to use the locations that users self-report in their profiles (free text).[2-4] Among the set of tweets collected for this study, 0.5% (1,735 of 358,194 tweets) included geo-tag information, while 70.1% (90,658 of 129,286 users) had some self-reported information about location in their user profiles.

Location inference methods were used to identify users located in Australia, Canada and United Kingdom. Nominatim,[5] a gazetteer, was used to translate the locations of Twitter users to identify users in the three countries. This information was taken from the tweet metadata (geo-tags) or user profile information (free text). Pre-processing steps for the user profile information included the removal of punctuation, numeric values, characters for non-English languages, and one-character words. Nominatim produces a score for the set of possible locations it returns, and a score of 0.4 was used as a threshold to avoid locations likely to be spurious (this threshold was determined through experiments in previous work). Where users included geo-tag information in the tweets they posted, the most frequent location was chosen (or the earliest where there were equally frequent and different locations used). Where users did not include geo-tag information, they were assigned to the location produced by Nominatim based on their user profile information.

2. Construction of the follower network

The social connections among the set of 16,789 users who posted about HPV vaccines and were located within Australia, Canada, and the UK were examined. The follower connections to and from each of the 16,789 users were collected through calls to the Twitter Application Program Interface (API), performed shortly after the first time each users posted a relevant tweet during the relevant time period. These data were used to construct the internal follower network by reconciling connections to and from each user to any other user in the set.

This study evaluated the proportions of international connections across the three countries, and examined the differences in the proportions of users who mostly post tweets expressing concerns about HPV vaccines relative to all other users. To do this, the ratios of follower connections of two types of users (those who express concerns in at least half of their relevant tweets versus all other users) in the three countries (Australia, Canada, and the UK), were compared.

3. Machine learning methods used to train and test the classification of the tweets

3.1. Pre-processing

The tweet texts were processed to construct features for our classifiers. No modifications were made on words that were hashtags (beginning with "#") and Twitter usernames (beginning with "@"). If a tweet text contained a website link (URLs), the domain name was stored. Standard data pre-processing including the removal of common English words,[6,7] and the removal of plurals and modifiers using Porter algorithm.[8] All numerical values were removed and all words were converted into lowercase. Each tweet was then transformed into a binary representation—a vector of length equal to the total number of unique features found across all tweets—with 1 marked for any feature in the tweet and 0 for all other features.

3.2. Supervised Machine Learning Training

Due to the large number of tweets collected in the period, a supervised machine learning approach was used to classify the tweets. This involved the manual labeling of a random sample of tweets, which were then used to train algorithms to identify similar patterns in the remaining tweets.[9,10] The data were classified in two stages to firstly identify tweets expressing concerns, and then to classify those concerns by type (Figure A1).

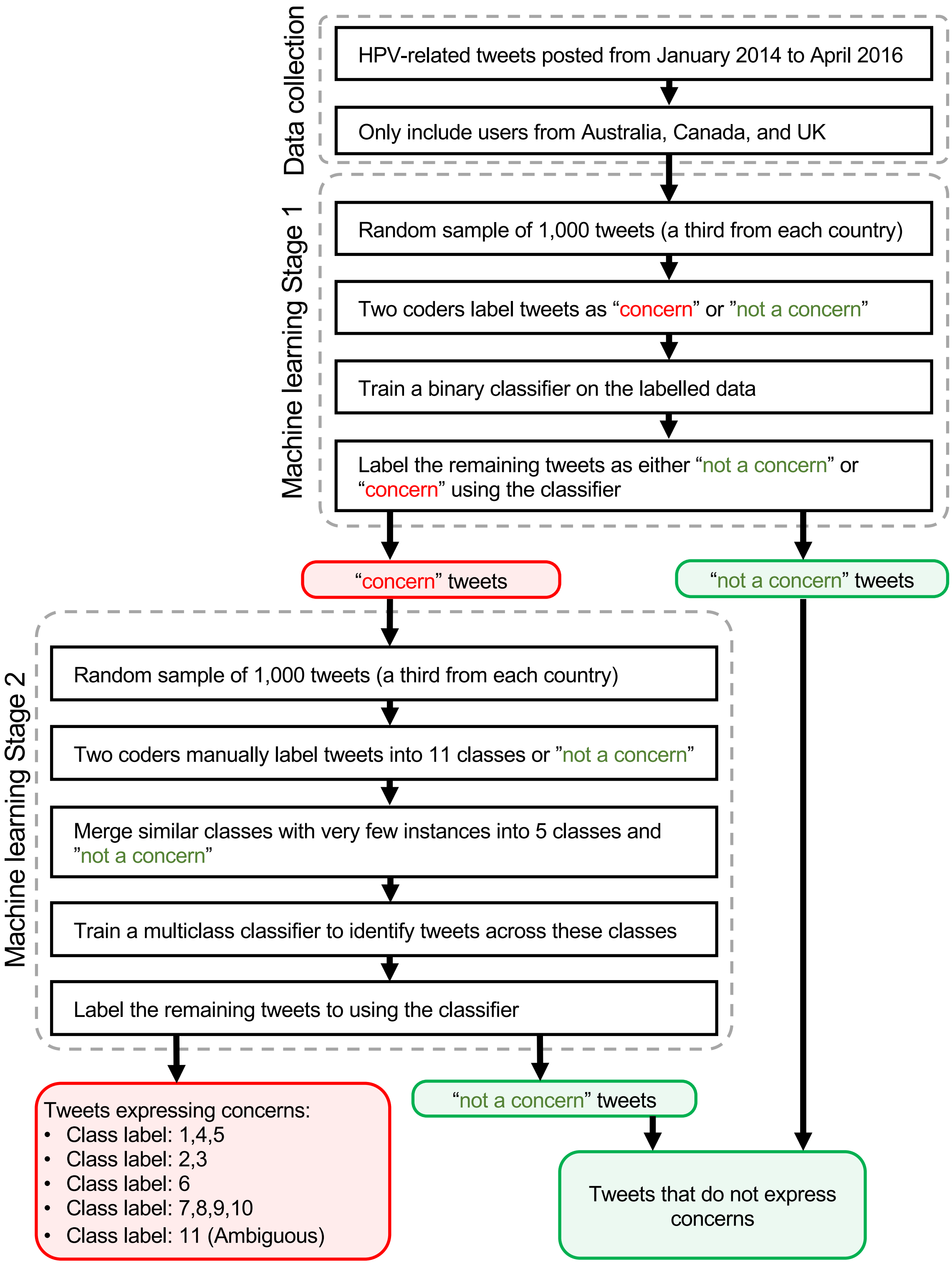


Figure A1. The design of the two-stage method for identifying classes of concerns in HPV vaccine tweets, where Stage 1 is the construction of a binary classifier and Stage 2 is the construction of a multiclass classifier

Classifiers for identifying any concern

The first stage of coding aimed to distinguish between tweets about HPV vaccines that expressed a concern versus tweets not expressing a concern. A random sample of 1,000 tweets were manually labeled as ‘concern’ or ‘non-concern’ by two investigators to form a training set from which to train a machine learning classifier. A separate set of 150 tweets was presented to the coders as a practice set for discussion prior to the independent labeling process. If the tone of the tweet was not immediately clear from the information provided within the tweet, the coders used links to webpages if they were available. There was strong agreement amongst the coders (92.3% agreement, Cohen’s $\varkappa$=0.81; 95% CI 0.77-0.85), and disagreements were resolved by discussion to produce the final training set.

A supervised binary classifier was trained using the manually labeled tweets to assign labels to the rest of the tweets in the data set. This study used a linear support vector machine (SVM) with stochastic gradient descent learning method to perform binary classification. The SVM method has been used widely for applications that deal with unbalanced and high dimensional data sets like those described here.[11-15] In this study, a random sample of 80% of the manually labeled tweets (the *training set*) was used to train and validate the classifier and the remaining 20% was used to test the performance of the classifiers (the *testing set*). The best parameters for the classifier were chosen using 10-fold cross validation using the training set. *K*-fold cross validation is a common method used to train a classifier in a prediction problem, where the training set is partitioned into *K* equal sized subsamples.[16] During the training process, the classifier is trained using all but one of the subsamples and validated on the remaining subsample, repeating the process *K* times. To avoid overfitting, the L2 regularization was used with 1,000 iterations during the training.[17,18]

Classifiers for identifying specific concerns

The second stage of coding aimed to distinguish different types of concern. The tweets classified as having expressed a concern about HPV vaccines in stage one were examined to distinguish the specific types of concerns. A random sample of 1,000 tweets was selected and a separate set of 150 tweets was used to pilot and test the scheme prior to the independent labeling process.

The categories for types of tweets expressing concerns were determined using an inductive and deductive procedure. Accordingly, the Health Belief Model (HBM) and additional concerns towards the HPV vaccine that have been identified in the literature were used to develop an initial coding scheme of 12 types of concerns (Table A1). These coding categories were discussed and agreed upon amongst the research team. There was good agreement in coding the random sample of 1,000 tweets (79.0% agreement, Cohen’s $\varkappa$=0.71; 95% CI 0.67-0.74), and any disagreements were resolved by discussion to produce the final training set. For example, the HBM factor of ‘self-efficacy’ was originally included in this coding scheme but deleted after team consultation as it overlapped with other groups during rounds of practice coding (i.e. ‘logistical barriers’).

Table A1. Number of coded tweets of each type of concern

Class label	Description	Number of tweets
1	Not beneficial	8
2	Perceived logistical challenges	18
3	Perceived harms	462
4	Not severe	3
5	Low susceptibility	1
6	Cues to action	141
7	Mistrust	90
8	Undermining religious principles	11
9	Undermining civil liberties	58
10	Additional concerns not otherwise specified	6
11	Tweet is ambiguous	18
12	No concern expressed	184
Total		1000

As can be seen in Table A1, some of the classes had fewer than 10 examples identified. Rare classes of concerns were merged based on similar themes to provide enough relevant examples to train and evaluate the performance of the multi-class classifier (Table A2).

Table A2. Number of coded tweets of each type of concern after merging the labels

Class label	Original class labels	Number of tweets
Unnecessary	1,4,5	12
Perceived barriers	2,3	480
Cues to action	6	141
Additional concerns	7,8,9,10	165
Ambiguous	11	18
Non-concern	12	184

In the second stage, the machine learning task was a multi-class classification. A *one-versus-rest* strategy was adopted where tweets from one class (type of concern) were treated as positive samples and all other tweets were treated as negative samples. A single linear support vector machine (SVM) with stochastic gradient descent learning method was trained as the classifier for each class and this was repeated for all types of concerns. The final label for each tweet was

assigned to the class for which there was the highest likelihood of it belonging to the positive class. Given the unbalanced nature of the labeled data (some classes have a large number of tweets while several others have a small number of tweets), a random sample of 65% of the labeled tweets (the *training set*) were used to train the classifiers and the remaining 35% of the labeled tweets (the *testing/holdout set*) were used to test the performance of the classifiers. The class weights were adjusted to be inversely proportional to the number of tweets in the classes in order to mitigate the influence effect of large classes during the training. The best parameters for the classifiers were chosen using the same approach as described above.

4. Proportional exposure to HPV vaccine related tweets

The total number of followers each of the users had at the time they posted their tweets was also used to measure the potential exposure to those tweets and the potential size of the audience for each class of concerns expressed by users within each country. To quantify the potential exposure to tweets by country and type of concern, the potential exposure to each tweet was defined by the number of followers that a user had at the time they posted a tweet about HPV vaccines.

Tweets expressing concerns tended to have smaller audiences compared with tweets not expressing concern about HPV vaccines (Tables A3 and A4). In Canada, tweets expressing concerns had a total potential exposure count of 3.75% (4.81 million of 128.4 million total potential exposures to tweets from users in Canada). In Australia, the proportion was 11.0% (3.25 million of 29.7 million total potential exposures to tweets from users in Australia), and in the UK, the proportion was 16.3% (21.3 million of 130.4 million total potential exposures to tweets from users in the UK).

The difference between the number of tweets and the relative sizes of the audiences show that expressions of concern about HPV vaccines were likely to have reached a smaller overall audience than would be expected given the number of tweets. Note that these numbers reflect the total number of exposures to each type of tweet rather than the total number of unique users who may have seen those tweets.

Table A3. The number and proportion of exposures to tweets classified as expressing concerns, by country

Country	Concern	Non-concern	Total
Australia (%)	3,254,528 (10.97%)	26,422,799 (89.03%)	29,677,327 (100%)
Canada (%)	4,810,618 (3.75%)	123,608,306 (96.25%)	128,418,924 (100%)
UK (%)	21,260,539 (16.30%)	109,182,707 (83.70%)	130,443,246 (100%)
Total	29,325685 (10.16%)	259,213,812 (89.84%)	288,539,497 (100%)

Table A4. The number and proportion of exposures to tweets posted by users, by country and type of concern

Group label	Australia (%)	Canada (%)	UK (%)	Total (%)
Unnecessary	7,508 (0.03%)	9,511 (0.01%)	53,384 (0.04%)	70,403 (0.02%)

Perceived barriers	1,950,348 (6.57%)	2,140,893 (1.67%)	9,912,306 (7.60%)	14,003,547(4.85%)
Cues to Action	287,686 (0.97%)	390,540 (0.30%)	803,537 (0.62%)	1,481,763 (0.51%)
Other concerns	595,832 (2.01%)	1,086,688 (0.85%)	1,978,772 (1.52%)	3,661,292 (1.27%)
Ambiguous	413,154 (1.39%)	1,182,986 (0.92%)	8,512,540 (6.53%)	10,108,680 (3.50%)
All non-concern	26,422,799 (89.03%)	123,608,306 (96.25%)	109,182,707 (83.70%)	259,213,812 (89.84%)
Total	29,677,327 (100%)	128,418,924(100%)	130,443,246 (100%)	288,539,497 (100%)

5. Performance of the classifiers

The binary classifier was designed to distinguish between tweets about HPV that expressed concerns from non-concerns. The binary classifier produced a precision of 90% and a recall of 90% (Table A5). In other words, approximately 1 in 10 tweets expressing a concern could have been misclassified as a non-concern tweet, and approximately 1 in 10 tweets not expressing a concern could have been misclassified as a tweet expressing a concern. Analyses reported should be interpreted in the context of this accuracy.

Table A5. Performance measures for the binary classifier within the testing/holdout set

Class label	Precision	Recall	F1 score	Number of tweets in the test set
Concern	0.90	0.97	0.93	143
Non-concern	0.89	0.74	0.81	57
Average/Total	0.90	0.90	0.90	200

The performance of the multi-class classifier varied relative to the number of instances available for training and testing in the labeled set of 1000 tweets (Table A6). The precision and recall were over 90% when identifying tweets from the 'cues to action' group, but a substantial proportion of tweets from other classes were misclassified as Class 11 (ambiguous tweets) when testing the classifier on the holdout. The performance results suggested that one could be reasonably confident about the proportions of tweets in 'perceived barriers' and 'cues to action' groups, but less confident about the proportions of tweets belonging to other classes.

Table A6. Performance measures for the multi-class classifier within the testing/holdout set

Class label	Precision	Recall	F1 score	Number of tweets in the test set
Not beneficial, not severe, & low susceptibility	0.50	0.14	0.22	7
Perceived barriers	0.81	0.74	0.77	180
Cues to Action	0.91	0.92	0.92	53
Other concerns	0.77	0.46	0.57	50

Ambiguous	0.03	0.4	0.06	5
Non-concern	0.45	0.31	0.37	55
Average/Total	0.74	0.64	0.68	350

6. Examination of the follower network

Examining the followers of users who expressed concerns about HPV vaccines, the results show that 34.7% of the followers of users expressing concerns were also sharing their concerns. In contrast, 8.3% of the followers of users who did not express concerns were users expressing concerns (Table A7).

Table A7. Aggregate percentages of followers for all countries and expression of concern

	Concern (%)				Non-concern (%)			
Internal network followers (aggregate follower count)	Australia	Canada	UK	All	Australia	Canada	UK	All
All concern (38,378)	4.5	17.8	12.4	34.7	9.9	18.4	37.0	65.3
All non-concern (464,251)	1.5	2.3	4.5	8.3	20.4	25.2	46.1	91.7
All users (502,629)	1.7	3.5	5.1	10.3	19.6	24.6	45.4	89.7

Examining the followers of users who expressed concerns about HPV vaccines, the results also show that these users were relatively well connected to users in other countries who also expressed concerns (Table A8). For example, 28.6% of the followers of Australian users expressing concerns were also users expressing concerns, and 52.4% of those followers were from Canada or the UK.

This type of social connection—between users from different countries—was disproportionately high between users expressing concerns about HPV vaccines, and this pattern was consistent across the three countries. These differences are also apparent in Figure 1 in the manuscript, where there is a higher density of users expressing concerns about HPV vaccines close to the boundaries between the clusters of users from Canada and the UK.

Table A8. Aggregate number and percentage of followers by country and expression of concern

	Concern (%)					Non-concern (%)				
Internal network followers (aggregate follower count)	Australia	Canada	UK	All	Proportion of international followers in the same concern group	Australia	Canada	UK	All	Proportion of international followers in the same concern group
Australian (102,894)	5.7	1.1	0.8	7.6		82.3	6.0	4.1	92.4	
-Concern (5,319)	13.6	8.3	6.7	28.6	52.4	54.4	6.5	10.5	74.4	23.8

-Non-concern (97,575)	5.3	0.7	0.4	6.4	17.2	83.8	6.0	3.7	93.6	10.4
Canadian (151,179)	0.9	9.6	1.6	12.0		6.6	71.8	9.6	88.0	
-Concern (14,656)	3.9	37.4	6.3	47.6	21.4	4.0	41.0	7.5	52.4	21.9
-Non-concern (136,523)	0.5	6.6	1.1	8.2	19.5	6.8	75.1	9.9	91.8	18.2
UK (248,556)	0.5	0.9	9.0	10.4		1.6	3.7	84.3	89.6	
-Concern (18,403)	2.3	5.0	18.8	26.1	28.0	1.8	3.9	63.1	73.9	8.3
-Non-concern (230,153)	0.4	0.5	8.2	9.1	9.9	1.6	3.7	85.6	90.9	5.8

References

1. Sloan L, Morgan J. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PloS one*. 2015;10(11):e0142209.
2. Elmongui HG, Morsy H, Mansour R. Inference models for twitter user's home location prediction. IEEE/ACS International Conference of Computer Systems and Applications (AICCSA); 2015.
3. Bo H, Paul C, Timothy B. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*. 2014;49:451-500.
4. Bo H, Paul C, Timothy B. A. Stacking-based approach to twitter user geolocation prediction. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistic; 2013.
5. Nominatim: OpenStreetMap Wiki. http://wiki.openstreetmap.org/wiki/Nominatim.
6. Lehmann J, Gonçalves B, Ramasco JJ, Cattuto C. *Dynamical Classes of Collective Attention in Twitter*. 2012.
7. Zhao WX, Jiang J, Weng J, et al. *Comparing Twitter and Traditional Media using Topic Models*. 2011.
8. Porter MF. *An algorithm for suffix stripping*. 1997.
9. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *Journal of medical Internet research*. 2015;17(6):e144.
10. Surian D, Nguyen QD, Kennedy G, Johnson M, Coiera E, Dunn GA. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *Journal of medical Internet research*. 2016;18(8):e232.
11. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in health technology and informatics*. 2015;216:761-765.
12. Mullen T, Collier N. *Sentiment analysis using support vector machines with diverse information sources*. Barcelona Spain,2004.
13. Pang B, Lee L, Vaithyanathan S. *Thumbs up? Sentiment classification using machine learning techniques*. Philadelphia United States,2002.

14. Sebastiani F. Machine learning in automated text categorization. *ACM Comput Surv*. 2002;34:1-47.
15. Joachims T. *Transductive inference for text classification using support vector machines*. Bled Slovenia,1999.
16. McLachlan GJ, Do K, Ambroise C. *Analyzing Microarray Gene Expression Data*. Wiley; 2004.
17. Coates A, Lee H, Ng A. An analysis of single-layer networks in unsupervised feature learning. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTAT); 2011.
18. Koshiba Y, Abe S. Comparison of L1 and L2 Support Vector Machines. Proceedings of the International Join Conference on Neural Networks; 2003.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time
		Case-control study—Report numbers in each exposure category, or summary measures of exposure
		Cross-sectional study—Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Comparing human papillomavirus vaccine concerns on Twitter: A cross-sectional study of users in Australia, Canada, and the United Kingdom

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016869.R2
Article Type:	Research
Date Submitted by the Author:	28-Jul-2017
Complete List of Authors:	Shapiro, Gilla; McGill University, Psychology ; University of Melbourne School of Population and Global Health, Centre for Health Policy Surian, Didi; Macquarie University, Centre for Health Informatics, Australian Institute of Health Innovation Dunn, Adam; Macquarie University, Centre for Health Informatics, Australian Institute of Health Innovation Perry, Ryan; University of Melbourne School of Population and Global Health, Centre for Health Policy Kelaher, Margaret; University of Melbourne School of Population and Global Health, Centre for Health Policy
Primary Subject Heading:	Communication
Secondary Subject Heading:	Sexual health, Public health, Oncology, Infectious diseases, Health informatics
Keywords:	Human papillomavirus, Vaccination, Twitter, Health Belief Model, Social media

Comparing human papillomavirus vaccine concerns on Twitter: A cross-sectional study of users in Australia, Canada, and the United Kingdom

Gilla K. Shapiro (corresponding author), PhD Candidate, Department of Psychology, McGill University, 2001 McGill College Avenue, Montreal, Quebec, H3A 1G1, Canada & Centre for Health Policy, School of Population and Global Health, University of Melbourne, Australia**;** Correspondence to: gilla.shapiro@mail.mcgill.ca

Didi Surian, Post-Doctoral Research Fellow, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 75 Talavera Road, Sydney, New South Wales, 2109, Australia

Adam Dunn, Senior Research Fellow**,** Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 75 Talavera Road, Sydney, New South Wales, 2109, Australia

Ryan Perry, Researcher, Centre for Health Policy, School of Population and Global Health, University of Melbourne, 235 Bouverie Street, Melbourne, Victoria, 3053, Australia

Margaret Kelaher, Professor, Head Evaluation and Implementation Science, Centre for Health Policy, School of Population and Global Health, University of Melbourne, 235 Bouverie Street, Melbourne, Victoria, 3053, Australia

Abstract

Objective: Opposition to human papillomavirus (HPV) vaccination is common on social media and has the potential to impact vaccine coverage. This study aims to conduct an international comparison of the proportions of tweets about HPV vaccines that express concerns, the types of concerns expressed, and the social connections among users posting about HPV vaccines in Australia, Canada and the United Kingdom.

Design: Using a cross-sectional design, an international comparison of English language tweets about HPV vaccines and social connections among Twitter users posting about HPV vaccines between January 2014 and April 2016 was conducted. The Health Belief Model (HBM), one of the most widely used theories in health psychology, was used as the basis for coding the types of HPV vaccine concerns expressed on Twitter.

Setting: The content of tweets and the social connections between users who posted tweets about HPV vaccines from Australia, Canada and the United Kingdom.

Population: 16,789 Twitter users who posted 43,852 tweets about HPV vaccines.

Main outcome measures: The proportions of tweets expressing concern, the type of concern expressed, and the proportions of local and international social connections between users.

Results: Tweets expressing concerns about HPV vaccines made up 14.9% of tweets in Canada, 19.4% in Australia, and 22.6% in the UK. The types of concerns expressed were similar across the three countries, with concerns related to 'perceived barriers' being the most common. Users expressing concerns about HPV vaccines in each of the three countries had a relatively high proportion of international followers also expressing concerns.

Conclusions: The proportions and types of HPV vaccine concerns expressed on Twitter were similar across the three countries. Twitter users who mostly expressed concerns about HPV

vaccines were better connected to international users who shared their concerns compared to users who did not express concerns about HPV vaccines.

Key words: Human papillomavirus; Vaccination; Twitter; Health Belief Model; Social media.

Strengths and limitations of this study

- This study conducted an international comparison investigating how HPV vaccination concerns are expressed on Twitter.
- Machine learning methods were used to identify and classify the proportion and types of concerns expressed in thousands of tweets.
- The analysis of social connections among Twitter users posting about HPV vaccines in three English-speaking countries (Australia, Canada, and the United Kingdom) revealed the potential for concerns to spread internationally.
- While Twitter is used by a substantial number of people, this study is not designed to allow for generalization to the general population.
- This study used follower networks to examine social connections but further research could use other social interactions to measure the spread and impact of negative attitudes about HPV vaccines.

Introduction

Human papillomavirus (HPV) is a prevalent sexually transmitted infection that can cause cancers and anogenital warts.[1-5] Since 2006, three prophylactic vaccines have been developed to protect adolescents from HPV-associated health problems.[6] Research has demonstrated that these vaccines are safe and effective in reducing HPV related infections, genital warts, and pre-cancers.[7-12] As a result, at least sixty-five countries have implemented HPV vaccination programs for females in their national immunization schedules.[11]

There is notable variation between countries' HPV vaccine programs and coverage rates. Australia's school-based vaccination program targets 12-13 year olds females (since 2007) and males (since 2013).[13] According to Australia's National HPV vaccination program register, 85.6% of females and 77% of males received the HPV vaccine (2015 data).[14,15] In Canada, all provinces and territories introduced school-based vaccination programs for 9-13 year old females (2007-2010), and six provinces also include boys in HPV vaccine programs (since 2013).[16] According to national parental surveys, 72.3% of females (2013 data) and less than 3% of males received the HPV vaccine (2014 data).[17-19] Lastly, the United Kingdom (UK) only provides a school-based vaccination program for 12-13 year old females (since 2008). According to Public Health England, 89.5% of females in the UK received the HPV vaccine (2015 data).[20] HPV vaccine coverage rates are lower than other child or adolescent vaccines in these countries national immunisation programs;[21-23] suboptimal coverage hinders cancer prevention efforts.[24]

The media has the potential to dramatically impact vaccine coverage through influencing parental awareness, perception, and attitudes.[25-29] Unconfirmed reports of adverse events associated with the HPV vaccine published in the media dramatically affected female HPV vaccine coverage in Japan and Colombia.[11,30,31] Many individuals use the internet and social media to access health information; however, these sources have been described as a risky platform that can rapidly amplify unbalanced, distorted or inaccurate information about vaccines.[25,32-34] For example, a study by Betsch et al. found that even 5 to 10 minutes of access to vaccine-critical websites negatively influenced individuals' risk perception and intentions to be vaccinated.[35] Similarly, Nan and Madden report that, compared to a control group, participants who were exposed to negative online blogs about HPV perceived the vaccine as less safe, held more negative attitudes, and reported a reduced intention to receive the vaccine.[36]

Previous research has evaluated the public discourse concerning HPV vaccination in newspapers,[37-40] online news,[41] comments to online news articles,[42] Facebook,[43] blogs or online forums,[36,44,45] and YouTube videos.[46,47] Twitter is a microblogging service, established in 2006, that has over 313 million users active monthly. Twitter is an important source of information regarding HPV and vaccine hesitancy,[32,48,49] and several studies have examined the representation of HPV vaccines on Twitter.[33,43,50-54] Though many of these studies analyse a limited number of HPV-related tweets, a few have used data mining and machine learning techniques to analyse a large number of tweets.[51,52,54,55] However, no research has conducted an international comparison to evaluate and compare how vaccination concerns are expressed across countries. Furthermore, no research has examined the domestic and international network connectedness of HPV vaccine concern expression.

The aim of this study was to explore the proportion of HPV vaccine concern on Twitter, examine the type of concern expressed in Australia, Canada, and the United Kingdom (UK), and investigate differences in the ways Twitter users connected locally and internationally.

Methods

Study overview

Tweets related to HPV vaccines during were collected from January 2014 to April 2016 in Australia, Canada, and the UK. These countries were selected because they are English-speaking countries, share a similar history and commonwealth membership, and their similarity in administering the HPV vaccination in schools. Data captured included information about users' locations, the text of the tweets, and information about social connections. To enable the classification of a large number of tweets, two stages of machine learning classifiers were constructed from a sample of tweets that were manually coded by two investigators.

Study data

Using a similar approach to previous studies that examined large number of tweets in communities of Twitter users posting about HPV vaccines,[54,56] the Twitter Search Application Programming Interface (API) was used to collect tweets in the English language about HPV vaccines from January 2014 to April, 2016. The search terms were "Gardasil", "Cervarix", "hpv AND vaccin*", and "cervical AND vaccin*". Information extracted from each tweet included the unique tweet identifier, tweet text, creation time, the identifier of the user posting the tweet, and geographical coordinates (if available). Without any restrictions applied to the locations of users, the entire dataset included 358,194 tweets (including retweets) by 129,286 users.

A gazetteer was used to transform the text provided by users into coordinates, and any users with self-reported locations that were located in coordinates in Australia, Canada, or the UK were included in this analysis (Supplemental Material, Section 1).

Other data that were used in the analyses included the set of social connections formed among the users who were included in the analyses. For each user, the Twitter Search API was used to collect the set of all follower relationships in which the user was involved, shortly after the first time the user posted a relevant tweet in the period. A network was then formed to include all users who tweeted about HPV vaccines from the three countries, and the follower relationships defined the social connections in an unweighted, directed network (Supplemental Material, Section 2).

The Macquarie University Human Research Ethics Committee (#5201401028) and the University of Melbourne's Research Ethics Board (#1647488.1) provided ethics approval for data collection and analysis.

Analysis

Supervised machine learning methods were used to classify the tweets in two stages; firstly to identify tweets that expressed any concern and secondly to classify specific types of concerns. In the first stage, 1000 tweets were sampled from the set of all tweets to manually label those that expressed concerns. In the second stage, 1000 tweets were sampled from the set of tweets that were estimated to be concerns to manually label them by type. The manually labelled tweets were used to train classifiers to label any tweet by the type of concern expressed (Supplemental Material, Section 3).

The categories for types of tweets expressing concerns were determined using an inductive and deductive procedure. The Health Belief Model (HBM), one of the most widely used theories in health psychology, was used as the basis for coding the types of HPV vaccine concerns expressed on Twitter.[57] The HBM has been used previously to evaluate the determinants of HPV vaccination and non-compliance by identifying perceived susceptibility to HPV, perceived severity of HPV, perceived benefits of HPV vaccination, perceived barriers of HPV vaccination (including tangible barriers such as logistical challenges and psychological barriers such as perceived harms of receiving the HPV vaccine), and cues to action (e.g. influences prompting HPV vaccine uptake such as information from health care provides, family, or friends).[47,57-59] To account for additional prominent concerns that were not captured by the model, the constructs were also informed by previous content analyses of media and social media related to the HPV vaccine, as well as literature on vaccine hesitancy.[28,33,37,41,42,45,47,60-68] The coding scheme was therefore extended to include mistrust, undermining of religious principles, undermining of civil liberties, additional concerns (not otherwise specified), and ambiguous tweets (Table 1). The coding scheme was used by two investigators (GS and RP) to code 12 types of concerns expressed in a second sample of 1000 tweets. After examining the proportions of different types of concerns in the sample and accuracy of the multi-class classifier (Supplemental Material, Sections 3 and 4), types of concerns were combined to improve the performance that could be achieved by the machine learning classifiers (Table 1). Combining categories was done based on conceptual similarity and trying to remain as true to the HBM as possible while attaining accuracy of the classifier.

The network of social connections formed by users who posted tweets about HPV vaccines was used to compare the proportions of local (within a country) and international (across countries) followers. The group of users for whom at least half of their relevant tweets were expressing concerns were assigned to one group (concern), and all other users were assigned to another (non-concern). The users were then also split by country, and the proportions of local and international followers were compared across groups.

Results

There were 129,286 Twitter users who posted at least one tweet about HPV vaccines during the period. The location inference method identified 2,792 (2.2%) of those users located in Australia, 7,237 (5.6%) located in Canada, and 6,760 (5.2%) users located in the UK (Table 2).

From the 16,789 users in the three countries, a total of 43,852 tweets about HPV vaccines were posted, of which 7,173 (16.4%) were from Australia, 18,927 (43.2%) were from Canada, and 17,752 (40.5%) were from the UK. This corresponded to an average of 2.57 tweets per user in Australia (range, 1-198), 2.61 tweets per user in Canada (range, 1-433), and 2.62 tweets per user in the UK (range, 1-501).

Expressions of concern

When labelling tweets that expressed concerns, the binary classifier (stage one) achieved a recall of 0.97 and a precision of 0.90. This indicates that the binary classifier missed 3% of tweets that were manually labelled as having expressed a concern and 10% of tweets it labelled as having expressed a concern were manually labelled otherwise. Because the multi-class classifier (at stage two) identified a proportion of these mislabelled tweets in the second round, the overall rate of error in stage one was within 5% of the correct proportion (Supplemental Material, Section 4).

The proportion of tweets posted about HPV vaccines from users in the three countries expressing concerns was 18.7% (8,215 of 43,852 tweets), but there were differences in these proportions across the three countries (Table 2). Canada had the lowest proportion of tweets expressing concerns at 14.9% (2,818 of 18,927 tweets), followed by Australia at 19.3% (1,388 of 7,173 tweets). The UK had the highest proportion of tweets expressing concerns at 22.6% (4,009 of 17,752 tweets). Tweets expressing concerns also tended to have smaller audiences compared with tweets not expressing concern about HPV vaccines (Supplemental Material, Section 3).

Types of concerns expressed

When identifying concerns related to cues to action the classifier respectively produced a precision of 0.81 and a recall of 0.74. For perceived barriers, the precision was 0.91 and recall was 0.92. The classifier was less reliable for the remainder of the concern groups because these types of concerns made up a much smaller proportions, resulting in imbalance in the data, which affects the performance that can be achieved by the classifiers (Supplemental Material, Section 5).

Tweets expressing concerns about perceived barriers comprised the largest type of concern by both the proportion of tweets expressing concerns (Table 3). The proportions of each group of concerns across the three countries were generally consistent.

Social connections among users

Among users from the three countries who posted about HPV vaccines, 18.2% (3,062 of 16,789) were labelled as having expressed concerns (at least half of the tweets about HPV vaccines they posted were labelled as having expressed a concern). The total number of follower connections among the set of 16,789 users was 502,629. Users from the three countries were disproportionately more likely to be followed by users from the same country, creating clusters of users by country (Figure 1). Furthermore, users who expressed concerns about the HPV vaccines appear to be more tightly connected within the United Kingdom, compared to either Australia or Canada. Figure 1 also highlights that users discussing HPV vaccines in the United Kingdom are more often connected to users in Australia and Canada than users in Australia and Canada are connected to each other.

To examine the proportion of followers of HPV vaccine tweets, Figure 2 examines “concern” and “non-concern” tweets for each of the three countries (to produce six groups represented as circles). Relative to users who did not express concern about HPV, users that did express concerns had a higher proportion of international followers who also expressed concerns

(Figure 2). Among UK users expressing concerns, 26.1% of followers also expressed concerns (compared to 9.1% of followers among UK users not expressing concerns). Also among UK users expressing concerns, 28.0% of their followers also expressed concerns and were from Australia or Canada, and 9.9% of their followers did not express concerns and were from Australia or Canada. In comparison, among UK users not expressing concerns, only 5.8% of their followers were users not expressing concerns and from Australia or Canada, and 8.3% of their followers were users expressing concerns and from Australia or Canada (Supplemental Material, Section 6). This pattern was consistent across each of the three countries. The results indicate that users who mostly expressed concerns were disproportionately well-connected to international users discussing HPV vaccines.

Discussion

This study found that in Australia, Canada, and the UK nearly one in five of the tweets about HPV vaccines were expressions of concern. Canadian Twitter users less often expressed concerns about HPV vaccines (14.9%) compared to Australia (19.3%) and the UK (22.6%) (Table 2). There was a general consistency in the proportions of specific concerns across the three countries, and the most common concerns (46%) were related to 'perceived barriers' (i.e. logistical challenges and psychological barriers such as perceived harms of receiving the HPV vaccine) (Table 3). The results demonstrated that users expressing concerns about HPV vaccines tended to be relatively well-connected to users discussing HPV vaccine concerns in other countries, especially between Canada and the UK.

Previous studies examining the representation of HPV vaccines on Twitter identified slightly higher proportions of negative tweets or tweets expressing concerns, but these studies captured different time periods and did not compare specific countries.[52,56] For example, a study of six months of Twitter data in the United States (between October 2013 and April

2014) found 25.1% of tweets were negative.[51] Though greater research is required, the balance of positive and negative content appears to vary by source whereby the majority of news content,[37] online comments (in response to news articles),[42] and tweets have been found to be positive; the majority of YouTube content has been found to be negative.[47] In examining the type of concern expressed about the HPV vaccine on other social media sites, researchers have also observed the predominance of perceived barrier (i.e. logistical challenges and psychological barriers such as perceived harms of receiving the HPV vaccine).[44,45,69] However, while the present research study found concerns about safety were most common on Twitter; other research found safety to be surpassed or similar in salience to other prevalent themes including conspiracies/search for truth, mistrust for health system, and promoting promiscuity.[44,45,69] Surian et al. analysed topics regarding HPV vaccines on Twitter and found individuals who posted about 'harms and conspiracies' posted more often than other users, suggesting that some users are actively seeking to introduce concerns about HPV vaccines into the public domain.[54] The predominance of HPV vaccine concerns about perceived barrier on Twitter indicates the importance of these concerns. It would be valuable to extend this work to examine differences in general vaccine concerns as well as compare concerns towards specific vaccines on Twitter.

International networking on Twitter suggests that vaccine related controversies in one country could reverberate around the world and impact vaccine coverage. Public health professionals and policymakers must therefore be able to monitor, rapidly identify, and react to such concerns (e.g. by providing evidence-based responses in real-time and strengthening their own international networks).[34,50,70] This research provides public health practitioners and policymakers with evidence that concerns about 'perceived barriers' on Twitter are widespread; effective communication campaigns could be designed and implemented to target this concern in locations where it is likely to have the greatest impact. However, it is important for further research to analyse results by type of sender. It would also be critical for future research to design and evaluate appropriate messaging of such a campaign so that this intervention does not 'backfire' and increase hesitancy.[71]

This study also found that Twitter users expressing HPV vaccine concerns tended to have higher proportions of international connections compared to those not expressing concerns. Given the international connection between twitter users who express concerns, public health organizations seeking to improve the uptake of HPV vaccines may benefit from tools that help them monitor the impact of vaccine scares on social media locally as well as in other countries in order to pre-empt and respond to misinformation. Greater research would be helpful to further investigate how public health organizations can monitor and intervene to address vaccine concerns. Such support could have been beneficial for Japan and Colombia when the media had a detrimental impact on HPV vaccine coverage.[11,30,31]

Similar to other studies in this area, our research did not measure whether the expression of concerns on Twitter led to changes in decision-making and coverage. Further research would be beneficial to assess the pathway of HPV vaccine concerns, and whether such concerns have a real-world impact (e.g. on vaccine coverage). In particular, future work should consider the relationship between the information about HPV vaccines that enters into the public discourse and the decision-making of individuals and populations. While Canada had the lowest proportion of tweets in which concerns were expressed of the three countries, it also has the lowest rate of HPV vaccine uptake.[14,15,18,19,72] Accordingly, we echo Gollust et al.'s recent call for greater experimental research designs to make causal assertions about the impact of the media on vaccine coverage.[73] Although some studies have begun to do so,[36,74] it would be helpful for future research to specifically evaluate the impact of Twitter messages and for moderating variables to also be evaluated.

Together with other studies on the representation of HPV vaccines in the media, our results suggest that it would be useful to monitor early indications of negative influence on attitudes and beliefs on social media. Two studies have independently examined responses on Twitter to specific controversial events including US Representative Michele Bachmann's claim that HPV vaccines could cause "mental retardation", and Katie Couric's television segment "HPV Vaccine Controversy" that aired on December 4, 2013 [53,75]. Mahoney *et al.* (2015) evaluated 200 social media posts before and after Bachmann's comments on the *Today Show* and found that though most media was positive in tone, compared to Google News, Twitter disseminated more positive HPV vaccine articles and also used more personal accounts as a reference source [53]. In contrast, using a random sample of 3,595 tweets, Bahk *et al.* (2016) found that most sentiment on Twitter towards HPV vaccines before Katie Couric's episode was negative, and while there was a decrease of negative sentiment immediately after the show aired, negative sentiment returned to baseline after two weeks [75]. Future research should also investigate how public health organizations should effectively intervene to curb misinformation or 'fake news' regarding HPV vaccination.

There were several limitations to this study. First, the findings are specific to Twitter, and while Twitter represents one of the largest populations of social media users, the results are not necessarily representative of the broader public discourse about HPV vaccines in news and online social media.[76-79] Twitter is an inherently biased representation of the broader population, and is skewed both in terms of age and socioeconomics.[80-83] Second, while the location inference method is a standard in the area,[84,85] the methods are imperfect.[86-89] Third, the study was limited to English-language tweets in three countries and evaluations of other countries and other languages may have yielded different results. It would be beneficial for future research to expand the focus of analysis to examine diverse countries, as well as conduct more nuanced regional explorations of a single country. Finally, using networks based on which users follow each other does not necessarily capture all of the interactions that occur online. While some argue that interaction with content (liking or retweeting) is a better measure of impact than followers,[90] others have argued that many users on Twitter are passive and do not interact with the content,[91] and as such followers may be a better indicator of impact. As this study examined follower networks, it would be helpful for future research to compare followers to different ways of interacting with content in order to better understand the impact of HPV vaccine tweets.

Conclusions

This study characterized the concerns about HPV vaccines expressed by Twitter users in three countries. The UK had the greatest proportion of tweets expressing concerns about HPV vaccines and Canada had the least, and the types of concerns expressed were relatively consistent across the three countries. Users who expressed concerns about HPV vaccines were generally more closely connected to users in other countries who also expressed concerns, suggesting that controversies and misinformation may be rapidly shared across international boundaries. This research could be used to design public health interventions that address concerns about the HPV vaccine on Twitter. In particular, this study suggests that methods for addressing vaccine concerns may benefit from targeting concerns about perceived barriers to vaccination (including logistical challenges and psychological barriers such as vaccine pain, safety, and side effects as a consequence of receiving the HPV vaccine). In addition, further coordination of public health agencies internationally may mitigate vaccine scares.

Figure 1 Legend: The follower network for Twitter users posting about HPV vaccines is coloured by country (Australia, green; Canada, red; UK, blue). Each node represents a user, and the node sizes are proportional to the number of followers within the user's network.

Nodes are positioned by heuristic to be closer to nodes with which they are better connected, as a way of illustrating the community structure. Darker coloured nodes indicate users for whom at least 50% of their relevant tweets expressed concerns.

Figure 2 Legend: The percentages of followers for all users by expression of concern from Australia, Canada and the UK. The circle represents a concern group of Twitter users, where the circle size is proportional to the number of users. The arrow represents user following direction. The number represents the percentage of followers, where the number in a circle represents the percentage of followers from the same concern group. Only values above 1.5% are shown, see Supplemental Material (Section 5) for all values.

Acknowledgement

Contributors GS, AD, and MK conceptualized and designed the study. DS and AD acquired the data. GS drafted the manuscript. All authors analysed and interpreted the data, critically revised the manuscript for important intellectual content, and approved the final version.

Funding: GS was supported by the CIHR Vanier Canada Graduate Scholarship and Queen Elizabeth II Diamond Jubilee Scholarship programs. AD is an investigator on a related NHMRC Project Grant (APP1128968). The funding source had no role in the design of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

Competing interests: None.

Ethics approval: Macquarie University Human Research Ethics Committee (#5201401028) and the University of Melbourne's Research Ethics Board (#1647488.1).

Data sharing statement: No additional data is available. Due to ethical approvals and the Twitter Terms of Service, the individual identifiers for Twitter posts and users cannot be provided with their manual or automatic labels.

References

1. Trottier H, Franco EL. The epidemiology of genital human papillomavirus infection. *Vaccine.* 2006;24(1):S1–15.
2. Tota JE, Chevarie-Davis M, Richardson LA, Devries M, Franco EL. Epidemiology and burden of HPV infection and related diseases: implications for prevention strategies. *Prev Med* 2011;53(1):S12-21.
3. Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *The Lancet Global Health.* 2016;4(9):e609–e616
4. Forman D, de Martel C, Lacey CJ, et al. Global burden of human papillomavirus and related diseases. *Vaccine.* Nov 20 2012;30 Suppl 5:F12-23.
5. National Advisory Committee on Immunization. Update On Human Papillomavirus Vaccines. *Canada Communicable Disease Report.* 2012;38.
6. Canadian Cancer Society. *Canadian Cancer Statistics 2016.* 2016.
7. Markowitz LE, Dunne EF, Saraiya M, et al. Human papillomavirus vaccination: recommendations of the Advisory Committee on Immunization Practices. *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control.* Aug 29 2014;63(Rr-05):1-30.
8. Ferris D, Samakoses R, Block SL, et al. Long-term study of a quadrivalent human papillomavirus vaccine. *Pediatrics.* Sep 2014;134(3):e657-665.
9. Garland SM. The Australian experience with the human papillomavirus vaccine. *Clin Ther* 2014;36:17-23.

10. Brotherton JM, Saville AM, May CL, Chappell G, Gertig DM. Human papillomavirus vaccination is changing the epidemiology of high-grade cervical lesions in Australia. *Cancer Causes Control.* 2015;26(6):953–954.

11. Brotherton JML, Bloem PJN. HPV Vaccination: Current Global Status. *Curr Obstet Gynecol Rep.* 2015;4:220-233.

12. Drolet M, Benard E, Boily MC, et al. Population-level impact and herd effects following human papillomavirus vaccination programmes: A systematic review and meta-analysis *Lancet Infect Dis.* 2015;15(5):565-580.

13. National Health Performance Authority. *Healthy Communities: HPV immunisation rates for girls in 2013 (In Focus)* 2015.

14. National HPV Vaccination Program Register. *National (Australia) HPV 3 dose vaccination coverage for females turning 15 years of age in 2015.* 2016.

15. National HPV Vaccination Program Register. National (Australia) HPV 3 dose vaccination coverage for males turning 15 years of age in 2015. 2016.

16. Shapiro Gilla K, Perez Samara, Rosberger Zeev. Including males in Canadian human papillomavirus vaccination programs: a policy analysis. *CMAJ.* 2016;188(12):881-886.

17. Samara Perez, Gilla K. Shapiro, Christopher A. Brown, Eve Dube, Gina Ogilvie, Rosberger Z. 'I didn't even know boys could get the vaccine': Parents' reasons for human papillomavirus (HPV) vaccination decision making for their sons. *Psycho-Oncology.* 2015.

18. Perez S, Tatar O, Shapiro GK, et al. Psychosocial determinants of parental human papillomavirus (HPV) vaccine decision-making for sons: Methodological challenges and initial results of a pan-Canadian longitudinal study. *BMC Public Health.* 2016;16(1):1223.

19. Gilbert NL, Gilmour H, Dubé E, Wilson SE, Laroche J. Estimates and determinants of HPV non-vaccination and vaccine refusal in girls 12 to 14 years of age in Canada: Results from the Childhood National Immunization Coverage Survey, 2013. *Human Vaccines & Immunotherapeutics.* 2016;12(6):1484-1490.

20. Bhavita Vishram, Lisa Byrne, Joanne White, Michael Edelstein. *Human Papillomavirus (HPV) vaccination coverage in adolescent females in England: 2014/15.* London, England: Public Health England;2015.

21. Australian Government Department of Health. Australian Immunisation Register: Current Data. In: Program IA, ed: Commonwealth of Australia; 2017.

22. Screening & Immunisations Team-NHS Digital. *NHS Immunisation Statistics England, 2015-16.* 2016.

23. Government of Canada. Vaccine coverage in Canadian children: Highlights from the 2013 childhood National Immunization Coverage Survey (cNICS). 2016.

24. Holman DM, Benard V, Roland KB, Watson M, Liddon N, Stokley S. Barriers to Human Papillomavirus Vaccination Among US Adolescents: A Systematic Review of the Literature. *JAMA pediatrics.* 2014; 168 (1):76-82.

25. Kumar D, Chandra R, Mathur M, Samdariya S, Kapoor N. Vaccine hesitancy: Understanding better to address better. *Isr J Health Policy Res.* 2016;5(2).

26. Mason BW, Donnelly PD. Impact of a local newspaper campaign on the uptake of the measles mumps and rubella vaccine. *J Epidemiol Community Health.* 2000;54:473-474.

27. Gangarosa EJ, Galazka AM, Wolfe CR, et al. Impact of anti-vaccine movements on pertussis control: the untold story. *Lancet.* 1998;351(9099):356-361.

28. Larson HJ, Jarrett C, Schulz WS, et al. Measuring vaccine hesitancy: The development of a survey tool. *Vaccine.* 2015.

29. Gollust SE, Attanasio L, Dempsey A, Benson AM, Fowler EF. Political and news media factors shaping public awareness of the HPV vaccine. *Womens Health Issues.* 2013;23(3):e143-151.

30. Hanley SJ, Yoshioka E, Ito Y, Kishi R. HPV vaccination crisis in Japan. *Lancet.* 2015;385(9987):2571.

31. El Tiempo. Caso de niñas del Carmen de Bolívar desplomó vacunación contra el VPH. February 14, 2015; http://www.eltiempo.com/estilo-de-vida/salud/vacuna-contra-el-papiloma-humano-cayo-58-puntos-porcentuales/15246061.

32. Gabarron E, Serrano JA, Wynn R, Lau AYS. Tweet content related to sexually transmitted diseases: no joking matter. *Journal of medical Internet research.* 2014;16(10):e228.

33. Britt RK, Hatten KN, Chappuis SO. Perceived behavioral control, intention to get vaccinated, and usage of online information about the human papillomavirus vaccine. *Health psychology and behavioral medicine.* 2014;2(1):52-65.

34. Amicizia D, Domnich A, Gasparini R, Bragazzi NL, Lai PL, Panatto D. An overview of current and potential use of information and communication technologies for immunization promotion among adolescents. *Human Vaccines & Immunotherapeutics* 2013;9(12):2634–2642.

35. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. *J Health Psychol.* 2010;15:446–455.

36. Nan X, Madden K. HPV vaccine information in the blogosphere: how positive and negative blogs influence vaccine-related risk perceptions, attitudes, and behavioral intentions. *Health communication.* 2012;27(8):829-836.

37. Perez S, Fedouruk C, Shapiro GK, Rosberger Z. Giving Boys a Shot: The HPV Vaccine's Portrayal in Canadian Newspapers. *Health communication.* 2016;31(12):1527-1538.

38. Robbins SC, Pang C, Leask J. Australian newspaper coverage of human papillomavirus vaccination, October 2006-December 2009. *Journal of health communication.* 2012;17(2):149-159.

39. Krieger JL, Katz ML, Eisenberg D, Heaner S, Sarge M, Jain P. Media coverage of cervical cancer and the HPV vaccine: implications for geographic health inequities. *Health expectations : an international journal of public participation in health care and health policy.* Sep 2013;16(3):e1-12.

40. Quintero JJ, Sionean C, Scott AM. Exploring the presentation of news information about the HPV vaccine: a content analysis of a representative sample of U.S. newspaper articles. *Health communication.* 2011;26(6):491-501.

41. Habel MA, Liddon N, Stryker JE. The HPV vaccine: A content analysis of online news stories. *J Womens Health.* 2009;18(3):401-407.

42. Feinberg Y, Pereira JA, Quach S, et al. Understanding Public Perceptions of the HPV Vaccination Based on Online Comments to Canadian News Articles. *PloS one.* 2015;10(6):e0129587.

43. Coloma PM, Becker B, Sturkenboom MC, van Mulligen EM, Kors JA. Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies. *Drug safety.* Oct 2015;38(10):921-930.

44. Penta MA, Baban A. Dangerous agent or saviour? HPV vaccine representations on online discussion forums in Romania. *International journal of behavioral medicine.* 2014;21(1):20-28.

45. Keelan J, Pavri V, Balakrishnan R, Wilson K. An analysis of the Human Papilloma Virus vaccine debate on MySpace blogs. *Vaccine.* Feb 10 2010;28(6):1535-1540.

46. Ache KA, Wallace LS. Human Papillomavirus vaccination coverage on YouTube. *Am J Prev Med* 2008;35(4):389-392.

47. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. *Health communication.* 2012;27:478–485.

48. Gunasekaran B, Jayasinghe Y, Fenner Y, et al. Knowledge of human papillomavirus and cervical cancer among young women recruited using a social networking site. *Sexually transmitted infections.* 2013;89(4):327-329.

49. McNeill A, Harris PR, Briggs P. Twitter Influence on UK Vaccination and Antiviral Uptake during the 2009 H1N1 Pandemic. *Frontiers in public health.* 2016;4:26.

50. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly available online tool facilitates real-time monitoring of vaccine conversations and sentiments. *Health Affairs.* 2016;35(2):341-347.

51. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *Journal of medical Internet research.* 2015;17(6):e144.

52. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in health technology and informatics.* 2015;216:761-765.

53. Mahoney LM, Tang T, Ji K, Ulrich-Schad J. The Digital Distribution of Public Health News Surrounding the Human Papillomavirus Vaccination: A Longitudinal Infodemiology Study. *JMIR Public Health Surveill.* 2015;1(39):e2.

54. Surian D, Nguyen QD, Kennedy G, Johnson M, Coiera E, Dunn GA. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *Journal of medical Internet research.* 08/29 2016;18(8):e232.

55. Salathé M, Vu DQ, Khandelwal S, Hunter DR. The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science.* 2013.

56. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *Journal of Medical Internet Research.* 2015;17(6):e144.

57. Rosenstock IM. Historical origins of the health belief model *Health Educ Monogr* 1974;2:328-335.

58. Donadiki EM, Jiménez-García R, Hernández-Barrera V, et al. Health Belief Model applied to non-compliance with HPV vaccine among female university students. *Public Health.* 2014;128(3):268-273.

59. Krawczyk A, Knäuper B, Gilca V, et al. Parents' decision-making about the human papillomavirus vaccine for their daughters: I. Quantitative results. *Hum Vaccin Immunother.* 2015;11(2):322-329.

60. Larson H, Leask J, Aggett S, Sevdalis N, Thomson A. A Multidisciplinary Research Agenda for Understanding Vaccine-Related Decisions. *Vaccines.* 2013;1(3):293-304.

61. Casciotti DM, Smith KC, Tsui A, Klassen AC. Discussions of adolescent sexuality in news media coverage of the HPV vaccine. *Journal of adolescence.* 2014;37(2):133-143.

62. Casciotti DM, Smith KC, Andon L, Vernick J, Tsui A, Klassen AC. Print news coverage of school-based human papillomavirus vaccine mandates. *The Journal of school health.* 2014;84(2):71-81.

63. Madden K, Nan X, Briones R, Waks L. Sorting through search results: a content analysis of HPV vaccine information online. *Vaccine.* 2012;30(25):3741-3746.

64. Nan X, Daily K. Biased assimilation and need for closure: Examining the effects of mixed blogs on vaccine-related beliefs. *Journal of health communication.* 2015;20(4):462-471.

65. Thomson A, Robinson K, Vallee-Tourangeau G. The 5As: A practical taxonomy for the determinants of vaccine uptake. *Vaccine.* 2016;34(8):1018-1024.

66. Wolfe RM, Sharp LK, Lipsky MS. Content and design attributes of antivaccination web sites. *JAMA.* 2002;287(24):3245-3248.

67. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: A content analysis. *JAMA.* 2007;298(21):2482-2484.

68. Abdelmutti N, Hoffman-Goetz L. Risk messages about HPV, cervical cancer, and the HPV vaccine Gardasil: A content analysis of Canadian and U.S. national newspaper articles. *Women & health.* Jul-Aug 2009;49(5):422-440.

69. Kata A. A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine.* 2010;28:1709-1716.

70. Larson HJ, Wilson R, Hanley S, Parys A, Paterson P. Tracking the global spread of vaccine sentiments: The global response to Japan's suspension of its HPV vaccine recommendation. *Hum Vaccin Immunother.* 2014;10(9):2543-2550.

71. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: A randomized trial. *Pediatrics.* 2014;133:e835-e842.

72. Vishram B, Byrne L, White J, Edelstein M. *Human Papillomavirus (HPV) vaccination coverage in adolescent females in England: 2014/15.* London, England: Public Health England;2015.

73. Gollust SE, LoRusso SM, Nagler RH, Fowlerc EF. Understanding the role of the news media in HPV vaccine uptake in the United States: Synthesis and commentary. *Hum Vaccin Immunother.* 2016;12(6):1430-1434.

74. Betsch C. Social media targeting of health messages: A promising approach for research and practice. *Hum Vaccin Immunother.* 2014;10(9):2636-2637.

75. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly Available Online Tool Facilitates Real-Time Monitoring Of Vaccine Conversations And Sentiments. *Health affairs (Project Hope).* Feb 2016;35(2):341-347.

76. Weeg C, Schwartz HA, Hill S, Merchant RM, Arango C, Ungar L. Using Twitter to Measure Public Discussion of Diseases: A Case Study. *JMIR Public Health Surveill.* 2015;1(1):e6.

77. Pew Research Center. *Social networking fact sheet.* http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/.

78. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in twitter: A quantitative analysis. *JMIR Public Health Surveill.* 2016;2(1):e1.

79. Paul MJ, Dredze M. You are what you tweet: Analyzing twitter for public health. *Association for the Advancement of Artificial Intelligence* 2011.

80. Sadah AS, Shahbazi M, Wiley TM ea. A study of the demographics of web-based health-related social media users. *Journal of medical Internet research.* 2015;17(8):e194.

81. Mislove A, Lehmann S, Ahn Y-Y, Onnela J-P, Rosenquist J.N. Understanding the demographics of Twitter users. Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM); 2011.

82. Sloan L, Morgan J, Burnap P, Williams M. Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data. . *PloS one.* 2015;10(3):e0115545.

83. Sloan L, Morgan J. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PloS one.* 2015;10(11):e0142209.

84. Kumar S, Morstatter F, Marshall G, Liu H, Nambiar U. Navigating Information Facets on Twitter (NIF-T). Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD); 2012.

85. Kounadi O, Lampoltshammer TJ, Groff E, Sitko I, Leitner M. Exploring twitter to analyze the public's reaction patterns to recently reported homicides in London. *PloS one.* 2015;10(3):e0121848.

86. Elmongui HG, Morsy H, Mansour R. Inference models for twitter user's home location prediction. IEEE/ACS International Conference of Computer Systems and Applications (AICCSA); 2015.

87. Bo H, Paul C, Timothy B. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research.* 2014;49:451-500.

88. Jurgens D. That's what friends are for: Inferring location in online social media platforms based on social relationships. . Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM); 2013.

89. Cheng Z, Caverlee J, Lee K. You are where you tweet: A content-based approach to geo-locating Twitter users. Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM); 2010.

90. Cha M, Haddadi H, Benevenuto F, Gummadi KP. Measuring user influence in Twitter: The million follower fallacy. Association for the Advancement of Artificial Intelligence 2010.

91. Romero DM, Galuba W, Asur S, Huberman BA. Influence and passivity in social media. In: Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, Michalis Vazirgiannis, eds. *Machine Learning and Knowledge Discovery in Databases* 2011:18-33.

Tables and Figures

Table 1. Coding scheme for the types of concerns expressed on Twitter

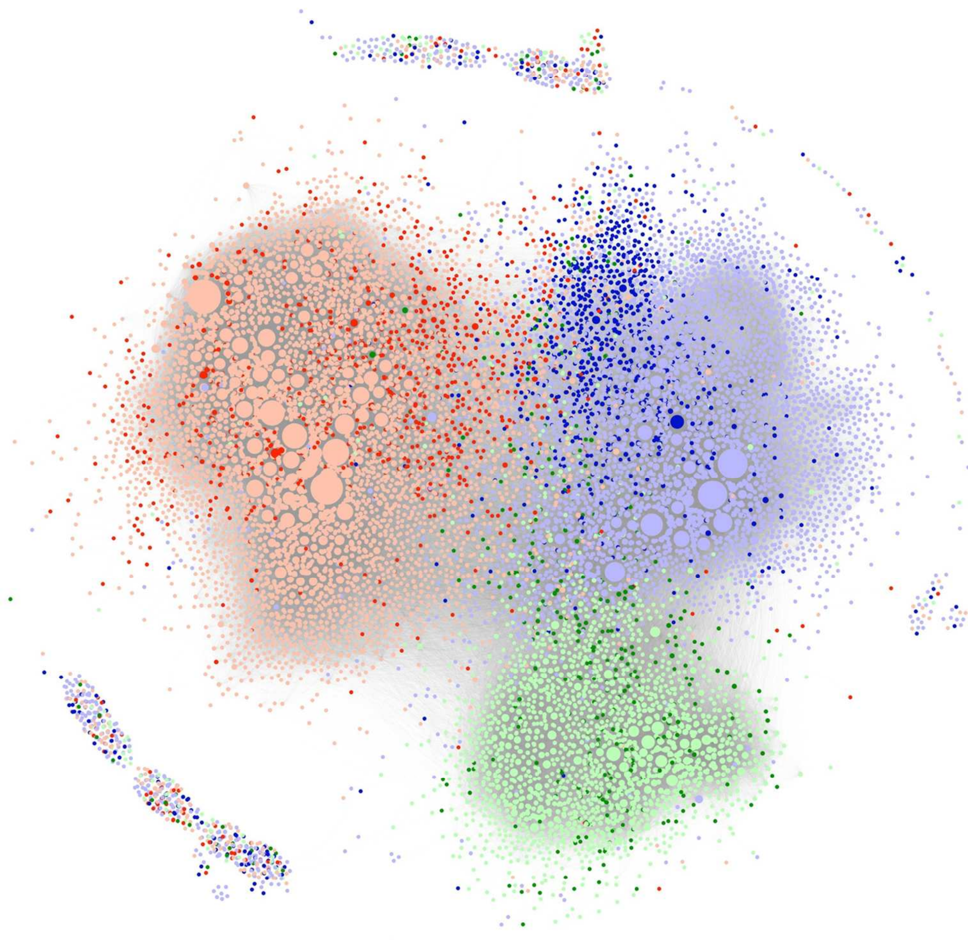
Original 12 types of concern	Combined concern group label	Example tweet
Not beneficial: stating that the HPV vaccine is not beneficial or useful (HBM construct 'perceived benefits')	Unnecessary	"Breaking Report HPV Cancers Rising In Spite of Vaccination URL #vaccination #guardasil #cervarix #HPV #cancer #fraud"
Perceived logistical challenges: stating logistical barriers such accessible or affordability challenges (HBM construct 'perceived barriers')	Perceived barriers	"Makes no sense that girls are covered for the HPV vaccine but I gotta pay $400 for it... #needstochange"
Perceived harms: stating concerns about the physical issues or harms as a result of receiving the HPV vaccine including pain, safety, or side effects (HBM construct 'perceived barriers')	Perceived barriers	"Mum still reckons im unwell like this all the time cos of the hpv vaccine. Thinking she might be right. Hpv vacc has loads of side effects"
Not severe: stating that HPV and/or its consequences are not severe (e.g. because it is common or clears up on its own) (HBM construct 'perceived severity')	Unnecessary	"The New Gardasil Is It Right For Your Daughter URL"
Low susceptibility: stating the HPV vaccine is unnecessary because there is a low likelihood of getting HPV and/or its consequences (HBM construct 'perceived susceptibility')	Unnecessary	"30 Facts you probably don\u2019t know about HPV and Gardasil...URL"
Cues to action: stating the influence of significant others guiding against receiving HPV vaccination (HBM construct 'cues to action')	Cues to action	"American College of Pediatricians warns about toxic effects of Gardasil vaccine"
Mistrust: stating a lack of confidence, mistrust, scepticism or belief in a HPV vaccine conspiracy	Additional concern	"Save dosh on the pharmaceutically lucrative, dubious Gardasil vaccine #qanda"
Undermining religious principles: stating concern that the HPV vaccine is inconsistent with religious principles	Additional concern	"…b. c. bishop, says chastity, not hpv vaccine, will keep girls healthy…"
Undermining civil liberties: stating concern about civil liberties (e.g. girls-only mandate, autonomy, who should be the decision maker for child vaccination, not being adequately consulted etc.)	Additional concern	"…had one dose of the gardasil at 17 after being bullied into it by my doctor, he basically told me I wasn't leaving without it"
Additional concerns not otherwise specified (e.g. belief in alternative medicine)	Additional concern	"…Our body does not need something NOT natural in our body to heal! The Gardasil/Vaccines were all in the..."
Tweet is ambiguous	Ambiguous	"…oh well if it's peer reviewed I'll give my son a gardasil shot."
No concern expressed	Non-concern	"Just saw a commercial that was like ask your doctor about Gardasil and I pumped my fist and shouted already did! because #sexualhealth"

Table 2. The total number of users and tweets from Australia, Canada, and the UK

Country	Number of users	Number of tweets	Tweets expressing concern	Tweets not expressing concern
Australia (%)	2,792 (16.6%)	7,173 (16.4%)	1,388 (19.4%)	5,785 (80.6%)
Canada (%)	7,237 (43.1%)	18,927 (43.2%)	2,818 (14.9%)	16,109 (85.1%)
UK (%)	6,760 (40.3%)	17,752 (40.5%)	4,009 (22.6%)	13,743 (77.4%)
Total	16,789 (100%)	43,852 (100%)	8,215 (18.7%)	35,637 (81.3%)

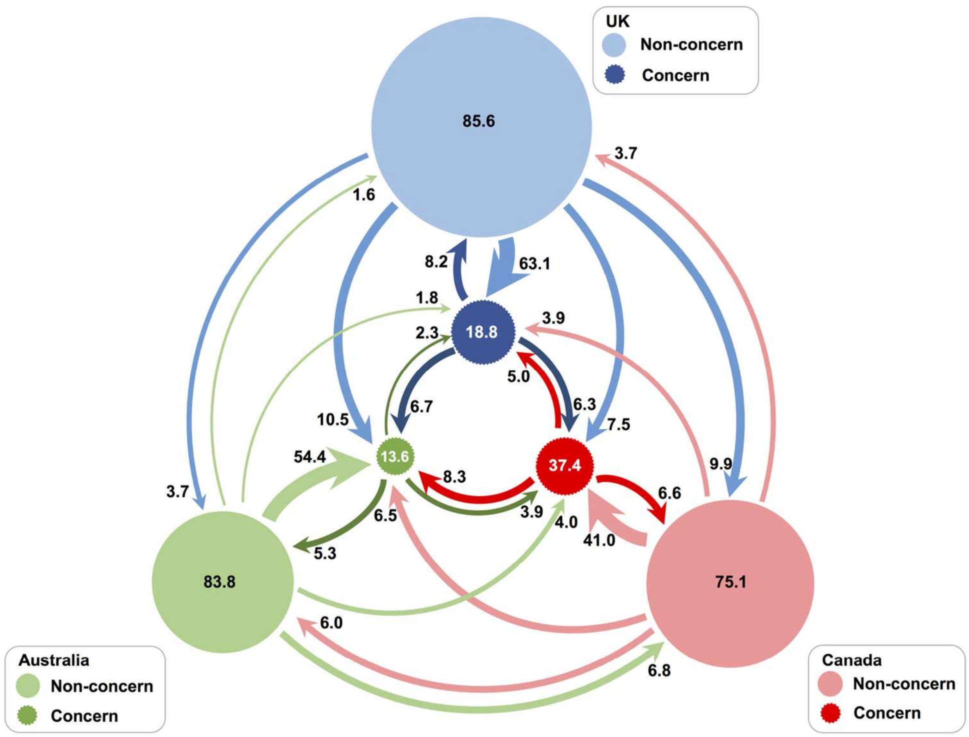
Table 3. Number of tweets, by country and concern type

Group label	Australia (%)	Canada (%)	UK (%)	Total (%)
Unnecessary	6 (0.39%)	13 (0.4%)	29 (0.6%)	48 (0.5%)
Perceived barriers	717 (47.08%)	1,368 (42.4%)	2,137 (48.0%)	4,222 (45.9%)
Cues to Action	157 (10.31%)	274 (8.5%)	306 (6.9%)	737 (8.0%)
Additional concerns	187 (12.28%)	469 (14.5%)	560 (12.6%)	1,216 (13.2%)
Ambiguous	321 (21.08%)	694 (21.5%)	977 (22.0%)	1,992 (21.7%)
Total concern	1,388 (100%)	2,818 (100 %)	4,009 (100%)	8,215 (100%)



The follower network for Twitter users posting about HPV vaccines is coloured by country (Australia, green; Canada, red; UK, blue). Each node represents a user, and the node sizes are proportional to the number of followers within the user's network. Nodes are positioned by heuristic to be closer to nodes with which they are better connected, as a way of illustrating the community structure. Darker coloured nodes indicate users for whom at least 50% of their relevant tweets expressed concerns.

127x127mm (300 x 300 DPI)



The percentages of followers for all users by expression of concern from Australia, Canada and the UK. The circle represents a concern group of Twitter users, where the circle size is proportional to the number of users. The arrow represents user following direction. The number represents the percentage of followers, where the number in a circle represents the percentage of followers from the same concern group. Only values above 1.5% are shown, see Supplemental Material (Section 5) for all values.

146x112mm (300 x 300 DPI)

Supplemental Material

The information described below provides greater detail of the methodology and results described in the manuscript.

1. Extracting and estimating the home locations of Twitter users

The use of 'geo-tags' on Twitter—where geographical coordinates are embedded in the metadata—is relatively rare and unevenly distributed, which makes geo-tags an unreliable way to determine the locations of users.[1] A common alternative is to use the locations that users self-report in their profiles (free text).[2-4] Among the set of tweets collected for this study, 0.5% (1,735 of 358,194 tweets) included geo-tag information, while 70.1% (90,658 of 129,286 users) had some self-reported information about location in their user profiles.

Location inference methods were used to identify users located in Australia, Canada and United Kingdom. Nominatim,[5] a gazetteer, was used to translate the locations of Twitter users to identify users in the three countries. This information was taken from the tweet metadata (geo-tags) or user profile information (free text). Pre-processing steps for the user profile information included the removal of punctuation, numeric values, characters for non-English languages, and one-character words. Nominatim produces a score for the set of possible locations it returns, and a score of 0.4 was used as a threshold to avoid locations likely to be spurious (this threshold was determined through experiments in previous work). Where users included geo-tag information in the tweets they posted, the most frequent location was chosen (or the earliest where there were equally frequent and different locations used). Where users did not include geo-tag information, they were assigned to the location produced by Nominatim based on their user profile information.

2. Construction of the follower network

The social connections among the set of 16,789 users who posted about HPV vaccines and were located within Australia, Canada, and the UK were examined. The follower connections to and from each of the 16,789 users were collected through calls to the Twitter Application Program Interface (API), performed shortly after the first time each users posted a relevant tweet during the relevant time period. These data were used to construct the internal follower network by reconciling connections to and from each user to any other user in the set.

This study evaluated the proportions of international connections across the three countries, and examined the differences in the proportions of users who mostly post tweets expressing concerns about HPV vaccines relative to all other users. To do this, the ratios of follower connections of two types of users (those who express concerns in at least half of their relevant tweets versus all other users) in the three countries (Australia, Canada, and the UK), were compared.

3. Machine learning methods used to train and test the classification of the tweets

3.1. Pre-processing

The tweet texts were processed to construct features for our classifiers. No modifications were made on words that were hashtags (beginning with "#") and Twitter usernames (beginning with "@"). If a tweet text contained a website link (URLs), the domain name was stored. Standard data pre-processing including the removal of common English words,[6,7] and the removal of plurals and modifiers using Porter algorithm.[8] All numerical values were removed and all words were converted into lowercase. Each tweet was then transformed into a binary representation—a vector of length equal to the total number of unique features found across all tweets—with 1 marked for any feature in the tweet and 0 for all other features.

3.2. Supervised Machine Learning Training

Due to the large number of tweets collected in the period, a supervised machine learning approach was used to classify the tweets. This involved the manual labeling of a random sample of tweets, which were then used to train algorithms to identify similar patterns in the remaining tweets.[9,10] The data were classified in two stages to firstly identify tweets expressing concerns, and then to classify those concerns by type (Figure A1).

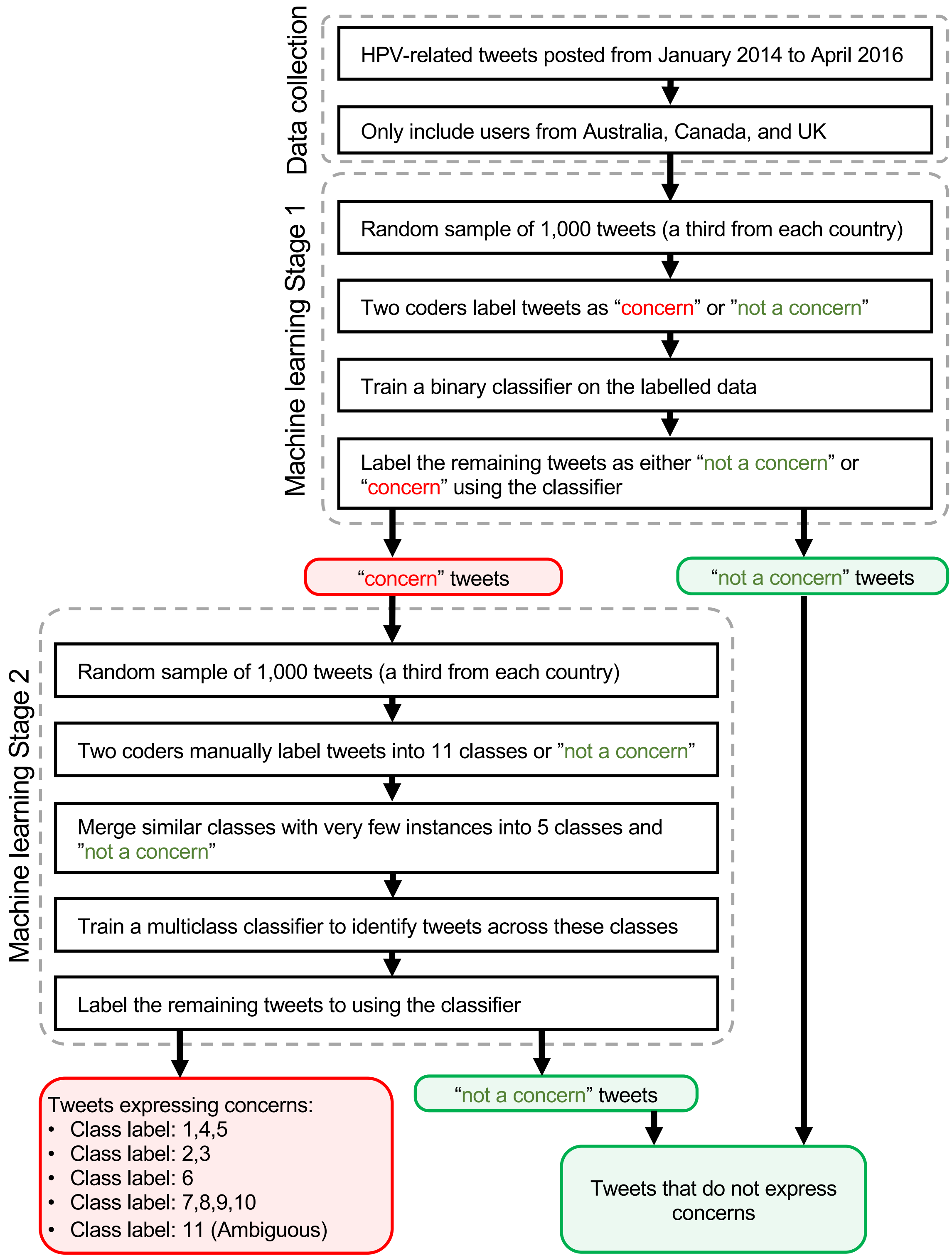


Figure A1. The design of the two-stage method for identifying classes of concerns in HPV vaccine tweets, where Stage 1 is the construction of a binary classifier and Stage 2 is the construction of a multiclass classifier

Classifiers for identifying any concern

The first stage of coding aimed to distinguish between tweets about HPV vaccines that expressed a concern versus tweets not expressing a concern. A random sample of 1,000 tweets were manually labeled as 'concern' or 'non-concern' by two investigators to form a training set from which to train a machine learning classifier. A separate set of 150 tweets was presented to the coders as a practice set for discussion prior to the independent labeling process. If the tone of the tweet was not immediately clear from the information provided within the tweet, the coders used links to webpages if they were available. There was strong agreement amongst the coders (92.3% agreement, Cohen's ϰ=0.81; 95% CI 0.77-0.85), and disagreements were resolved by discussion to produce the final training set.

A supervised binary classifier was trained using the manually labeled tweets to assign labels to the rest of the tweets in the data set. This study used a linear support vector machine (SVM) with stochastic gradient descent learning method to perform binary classification. The SVM method has been used widely for applications that deal with unbalanced and high dimensional data sets like those described here.[11-15] In this study, a random sample of 80% of the manually labeled tweets (the *training set*) was used to train and validate the classifier and the remaining 20% was used to test the performance of the classifiers (the *testing set*). The best parameters for the classifier were chosen using 10-fold cross validation using the training set. *K*-fold cross validation is a common method used to train a classifier in a prediction problem, where the training set is partitioned into *K* equal sized subsamples.[16] During the training process, the classifier is trained using all but one of the subsamples and validated on the remaining subsample, repeating the process *K* times. To avoid overfitting, the L2 regularization was used with 1,000 iterations during the training.[17,18]

Classifiers for identifying specific concerns

The second stage of coding aimed to distinguish different types of concern. The tweets classified as having expressed a concern about HPV vaccines in stage one were examined to distinguish the specific types of concerns. A random sample of 1,000 tweets was selected and a separate set of 150 tweets was used to pilot and test the scheme prior to the independent labeling process.

The categories for types of tweets expressing concerns were determined using an inductive and deductive procedure. Accordingly, the Health Belief Model (HBM) and additional concerns towards the HPV vaccine that have been identified in the literature were used to develop an initial coding scheme of 12 types of concerns (Table A1). These coding categories were discussed and agreed upon amongst the research team. There was good agreement in coding the random sample of 1,000 tweets (79.0% agreement, Cohen's ϰ=0.71; 95% CI 0.67-0.74), and any disagreements were resolved by discussion to produce the final training set. For example, the HBM factor of 'self-efficacy' was originally included in this coding scheme but deleted after team consultation as it overlapped with other groups during rounds of practice coding (i.e. 'logistical barriers').

Table A1. Number of coded tweets of each type of concern

Class label	Description	Number of tweets
1	Not beneficial	8
2	Perceived logistical challenges	18
3	Perceived harms	462
4	Not severe	3
5	Low susceptibility	1
6	Cues to action	141
7	Mistrust	90
8	Undermining religious principles	11
9	Undermining civil liberties	58
10	Additional concerns not otherwise specified	6
11	Tweet is ambiguous	18
12	No concern expressed	184
Total		1000

As can be seen in Table A1, some of the classes had fewer than 10 examples identified. Rare classes of concerns were merged based on similar themes to provide enough relevant examples to train and evaluate the performance of the multi-class classifier (Table A2).

Table A2. Number of coded tweets of each type of concern after merging the labels

Class label	Original class labels	Number of tweets
Unnecessary	1,4,5	12
Perceived barriers	2,3	480
Cues to action	6	141
Additional concerns	7,8,9,10	165
Ambiguous	11	18
Non-concern	12	184

In the second stage, the machine learning task was a multi-class classification. A *one-versus-rest* strategy was adopted where tweets from one class (type of concern) were treated as positive samples and all other tweets were treated as negative samples. A single linear support vector machine (SVM) with stochastic gradient descent learning method was trained as the classifier for each class and this was repeated for all types of concerns. The final label for each tweet was

assigned to the class for which there was the highest likelihood of it belonging to the positive class. Given the unbalanced nature of the labeled data (some classes have a large number of tweets while several others have a small number of tweets), a random sample of 65% of the labeled tweets (the *training set*) were used to train the classifiers and the remaining 35% of the labeled tweets (the *testing/holdout set*) were used to test the performance of the classifiers. The class weights were adjusted to be inversely proportional to the number of tweets in the classes in order to mitigate the influence effect of large classes during the training. The best parameters for the classifiers were chosen using the same approach as described above.

4. Proportional exposure to HPV vaccine related tweets

The total number of followers each of the users had at the time they posted their tweets was also used to measure the potential exposure to those tweets and the potential size of the audience for each class of concerns expressed by users within each country. To quantify the potential exposure to tweets by country and type of concern, the potential exposure to each tweet was defined by the number of followers that a user had at the time they posted a tweet about HPV vaccines.

Tweets expressing concerns tended to have smaller audiences compared with tweets not expressing concern about HPV vaccines (Tables A3 and A4). In Canada, tweets expressing concerns had a total potential exposure count of 3.75% (4.81 million of 128.4 million total potential exposures to tweets from users in Canada). In Australia, the proportion was 11.0% (3.25 million of 29.7 million total potential exposures to tweets from users in Australia), and in the UK, the proportion was 16.3% (21.3 million of 130.4 million total potential exposures to tweets from users in the UK).

The difference between the number of tweets and the relative sizes of the audiences show that expressions of concern about HPV vaccines were likely to have reached a smaller overall audience than would be expected given the number of tweets. Note that these numbers reflect the total number of exposures to each type of tweet rather than the total number of unique users who may have seen those tweets.

Table A3. The number and proportion of exposures to tweets classified as expressing concerns, by country

Country	Concern	Non-concern	Total
Australia (%)	3,254,528 (10.97%)	26,422,799 (89.03%)	29,677,327 (100%)
Canada (%)	4,810,618 (3.75%)	123,608,306 (96.25%)	128,418,924 (100%)
UK (%)	21,260,539 (16.30%)	109,182,707 (83.70%)	130,443,246 (100%)
Total	29,325685 (10.16%)	259,213,812 (89.84%)	288,539,497 (100%)

Table A4. The number and proportion of exposures to tweets posted by users, by country and type of concern

Group label	Australia (%)	Canada (%)	UK (%)	Total (%)
Unnecessary	7,508 (0.03%)	9,511 (0.01%)	53,384 (0.04%)	70,403 (0.02%)

Perceived barriers	1,950,348 (6.57%)	2,140,893 (1.67%)	9,912,306 (7.60%)	14,003,547(4.85%)
Cues to Action	287,686 (0.97%)	390,540 (0.30%)	803,537 (0.62%)	1,481,763 (0.51%)
Other concerns	595,832 (2.01%)	1,086,688 (0.85%)	1,978,772 (1.52%)	3,661,292 (1.27%)
Ambiguous	413,154 (1.39%)	1,182,986 (0.92%)	8,512,540 (6.53%)	10,108,680 (3.50%)
All non-concern	26,422,799 (89.03%)	123,608,306 (96.25%)	109,182,707 (83.70%)	259,213,812 (89.84%)
Total	29,677,327 (100%)	128,418,924(100%)	130,443,246 (100%)	288,539,497 (100%)

5. Performance of the classifiers

The binary classifier was designed to distinguish between tweets about HPV that expressed concerns from non-concerns. The binary classifier produced a precision of 90% and a recall of 90% (Table A5). In other words, approximately 1 in 10 tweets expressing a concern could have been misclassified as a non-concern tweet, and approximately 1 in 10 tweets not expressing a concern could have been misclassified as a tweet expressing a concern. Analyses reported should be interpreted in the context of this accuracy.

Table A5. Performance measures for the binary classifier within the testing/holdout set

Class label	Precision	Recall	F1 score	Number of tweets in the test set
Concern	0.90	0.97	0.93	143
Non-concern	0.89	0.74	0.81	57
Average/Total	0.90	0.90	0.90	200

The performance of the multi-class classifier varied relative to the number of instances available for training and testing in the labeled set of 1000 tweets (Table A6). The precision and recall were over 90% when identifying tweets from the 'cues to action' group, but a substantial proportion of tweets from other classes were misclassified as Class 11 (ambiguous tweets) when testing the classifier on the holdout. The performance results suggested that one could be reasonably confident about the proportions of tweets in 'perceived barriers' and 'cues to action' groups, but less confident about the proportions of tweets belonging to other classes.

Table A6. Performance measures for the multi-class classifier within the testing/holdout set

Class label	Precision	Recall	F1 score	Number of tweets in the test set
Not beneficial, not severe, & low susceptibility	0.50	0.14	0.22	7
Perceived barriers	0.81	0.74	0.77	180
Cues to Action	0.91	0.92	0.92	53
Other concerns	0.77	0.46	0.57	50

Ambiguous	0.03	0.4	0.06	5
Non-concern	0.45	0.31	0.37	55
Average/Total	0.74	0.64	0.68	350

6. Examination of the follower network

Examining the followers of users who expressed concerns about HPV vaccines, the results show that 34.7% of the followers of users expressing concerns were also sharing their concerns. In contrast, 8.3% of the followers of users who did not express concerns were users expressing concerns (Table A7).

Table A7. Aggregate percentages of followers for all countries and expression of concern

	Concern (%)				Non-concern (%)			
Internal network followers (aggregate follower count)	Australia	Canada	UK	All	Australia	Canada	UK	All
All concern (38,378)	4.5	17.8	12.4	34.7	9.9	18.4	37.0	65.3
All non-concern (464,251)	1.5	2.3	4.5	8.3	20.4	25.2	46.1	91.7
All users (502,629)	1.7	3.5	5.1	10.3	19.6	24.6	45.4	89.7

Examining the followers of users who expressed concerns about HPV vaccines, the results also show that these users were relatively well connected to users in other countries who also expressed concerns (Table A8). For example, 28.6% of the followers of Australian users expressing concerns were also users expressing concerns, and 52.4% of those followers were from Canada or the UK.

This type of social connection—between users from different countries—was disproportionately high between users expressing concerns about HPV vaccines, and this pattern was consistent across the three countries. These differences are also apparent in Figure 1 in the manuscript, where there is a higher density of users expressing concerns about HPV vaccines close to the boundaries between the clusters of users from Canada and the UK.

Table A8. Aggregate number and percentage of followers by country and expression of concern

	Concern (%)					Non-concern (%)				
Internal network followers (aggregate follower count)	Australia	Canada	UK	All	Proportion of international followers in the same concern group	Australia	Canada	UK	All	Proportion of international followers in the same concern group
Australian (102,894)	5.7	1.1	0.8	7.6		82.3	6.0	4.1	92.4	
-Concern (5,319)	13.6	8.3	6.7	28.6	52.4	54.4	6.5	10.5	74.4	23.8

-Non-concern (97,575)	5.3	0.7	0.4	6.4	17.2	83.8	6.0	3.7	93.6	10.4
Canadian (151,179)	0.9	9.6	1.6	12.0		6.6	71.8	9.6	88.0	
-Concern (14,656)	3.9	37.4	6.3	47.6	21.4	4.0	41.0	7.5	52.4	21.9
-Non-concern (136,523)	0.5	6.6	1.1	8.2	19.5	6.8	75.1	9.9	91.8	18.2
UK (248,556)	0.5	0.9	9.0	10.4		1.6	3.7	84.3	89.6	
-Concern (18,403)	2.3	5.0	18.8	26.1	28.0	1.8	3.9	63.1	73.9	8.3
-Non-concern (230,153)	0.4	0.5	8.2	9.1	9.9	1.6	3.7	85.6	90.9	5.8

References

1. Sloan L, Morgan J. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PloS one*. 2015;10(11):e0142209.
2. Elmongui HG, Morsy H, Mansour R. Inference models for twitter user's home location prediction. IEEE/ACS International Conference of Computer Systems and Applications (AICCSA); 2015.
3. Bo H, Paul C, Timothy B. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*. 2014;49:451-500.
4. Bo H, Paul C, Timothy B. A. Stacking-based approach to twitter user geolocation prediction. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistic; 2013.
5. Nominatim: OpenStreetMap Wiki. http://wiki.openstreetmap.org/wiki/Nominatim.
6. Lehmann J, Gonçalves B, Ramasco JJ, Cattuto C. *Dynamical Classes of Collective Attention in Twitter*. 2012.
7. Zhao WX, Jiang J, Weng J, et al. *Comparing Twitter and Traditional Media using Topic Models*. 2011.
8. Porter MF. *An algorithm for suffix stripping*. 1997.
9. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *Journal of medical Internet research*. 2015;17(6):e144.
10. Surian D, Nguyen QD, Kennedy G, Johnson M, Coiera E, Dunn GA. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *Journal of medical Internet research*. 2016;18(8):e232.
11. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in health technology and informatics*. 2015;216:761-765.
12. Mullen T, Collier N. *Sentiment analysis using support vector machines with diverse information sources*. Barcelona Spain,2004.
13. Pang B, Lee L, Vaithyanathan S. *Thumbs up? Sentiment classification using machine learning techniques*. Philadelphia United States,2002.

14. Sebastiani F. Machine learning in automated text categorization. *ACM Comput Surv*. 2002;34:1-47.
15. Joachims T. *Transductive inference for text classification using support vector machines*. Bled Slovenia,1999.
16. McLachlan GJ, Do K, Ambroise C. *Analyzing Microarray Gene Expression Data*. Wiley; 2004.
17. Coates A, Lee H, Ng A. An analysis of single-layer networks in unsupervised feature learning. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTAT); 2011.
18. Koshiba Y, Abe S. Comparison of L1 and L2 Support Vector Machines. Proceedings of the International Join Conference on Neural Networks; 2003.

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*
Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1-2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any pre-specified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	7-8
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	8-9
Bias	9	Describe any efforts to address potential sources of bias	15
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-9
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	N/A
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	9
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9-11
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	Figure A1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9-10
		(b) Indicate number of participants with missing data for each variable of interest	N/A
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	N/A
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	N/A
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	N/A
		Cross-sectional study—Report numbers of outcome events or summary measures	10-12
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	10-12
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10-12
Discussion			
Key results	18	Summarise key results with reference to study objectives	12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15-16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	16
Generalisability	21	Discuss the generalisability (external validity) of the study results	16
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	18

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.